

INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTING

Volume 04 Issue 02



International Journal of Research in Computing (IJRC)

Volume 04 Issue 02

July 2025

ISSN 2820-2139

© 2025 Faculty of Computing, General Sir John Kotelawala Defence University, Sri Lanka.

All rights reserved.

No part of this publication may be reproduced or quoted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without permission in writing from the Faculty of Computing of General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka.

Published by

Faculty of Computing, General Sir John Kotelawala Defence University,
Ratmalana, Sri Lanka

Tel: +94-11-2635268

E-Mail: editor@ijrcom.org & editorijrc@kdu.ac.lk

Website: <http://ijrcom.org/>

EDITORIAL COMMITTEE

CHIEF ADVISORS

Dr. HL Premarathne

Senior Lecturer (Retired)

School of Computing, University of Colombo, Sri Lanka

Dr. LP Kalansooriya

Dean

Faculty of Computing

General Sir John Kotelawala Defence University, Sri Lanka

EDITOR IN CHIEF

Dr. ADAI Gunasekara

Senior Lecturer

Department of Computer Science, Faculty of Computing

General Sir John Kotelawala Defence University, Sri Lanka

ASSOCIATE EDITORS IN CHIEF

Prof. TL Weerawardane

Dean/Professor of Electronics and Telecommunication

Faculty of Engineering

General Sir John Kotelawala Defence University, Sri Lanka

Dr. (Mrs). HRWP Gunathilake

Senior Lecturer

Department of Computer Science, Faculty of Computing

General Sir John Kotelawala Defence University, Sri Lanka

Dr. (Mrs). N Wedasinghe

Senior Lecturer

Department of Information Technology, Faculty of Computing

General Sir John Kotelawala Defence University, Sri Lanka

MEMBERS OF THE EDITORIAL BOARD

Prof. Yukun Bao

Deputy Director of Center for Modern Information Systems
Huazhong University of Science & Technology, China

Prof. R.Hoque

Professor
Law at the University of Dhaka, Bangladesh

Prof. Shamim Kaiser

Professor
Institute of Information Technology
Jahangirnagar University, Bangladesh

Dr. Attaphongse Taparugssanagorn

Associate Professor
School of Engineering and Technology
Asian Institute of Technology, Thailand

Snr.Prof. AS Karunananda

Senior professor
Department of Computational Mathematics, Faculty of Information Technology
University of Moratuwa, Sri Lanka

Prof. Prasad Jayaweera

Head/Professor of Computer Science
Department of Computer Science, Faculty of Applied Sciences
University of Sri Jayawardhanapura, Sri Lanka

Assoc. Prof. Anuja Dharmaratne

Associate Head (Education) School of IT
Monash University, Australia

Dr. Romuald Jolivot

Research Scholar
School of Engineering
Bangkok University, Thailand

Dr. MB Dissanayake

Senior Lecturer
Department of Electrical and Electronic Engineering
University of Peradeniya, Sri Lanka

Dr. APR Wickramarachchi

Senior Lecturer

Department of Industrial Management

University of Kelaniya, Sri Lanka

EDITORIAL ASSISTANTS

Ms. DVDS Abeysinghe

Lecturer (Probationary)

Department of Computer Science

Faculty of Computing

General Sir John Kotelawala Defence University, Sri Lanka

Ms. KD Madhubashani

Instructor

Department of Computational Mathematics

Faculty of Computing

General Sir John Kotelawala Defence University, Sri Lanka

CONTENTS

Systematic Review on AI in Gender Bias Detection and Mitigation in Education and Workplaces	(1 - 11)
<i>D Deckker and S Sumanasekara</i>	
Augmented Reality (AR) and Virtual Reality (VR) in Education: A Comprehensive Review	(12 - 24)
<i>KMHL Konara, GK Dilani, TMHC Peiris, RMR Dileka, TP Rathnayaka, WMCJT Kithulwatta, RMD Jayathilake, YNS Wijewardana, HMCC Somarathna and RMKT Rathnayake</i>	
Artificial Intelligence in Smart Cities and Urban Mobility: A Systematic Literature Review	(25 - 36)
<i>K Luxshi, RMKT Rathnayaka, DMKN Seneviratna and WMCJT Kithulwatta</i>	
Innovative ECG Classification Approach Utilizing a Transfer Learning-Driven Ensemble Architecture	(37 - 43)
<i>HMLS Kumari</i>	
AI-Driven Disaster Prediction and Early Warning Systems: A Systematic Literature Review	(44 - 55)
<i>K Luxshi</i>	
Speech Emotion Recognition with Hybrid CNN- LSTM and Transformers Models: Evaluating the Hybrid Model Using Grad-CAM	(56 – 66)
<i>HMLS Kumari, HMNS Kumari and UMMPK Nawarathne</i>	

Systematic Review on AI in Gender Bias Detection and Mitigation in Education and Workplaces

D Deckker^{1#}, S Sumanasekara²

¹Wrexham University, United Kingdom

²University of Gloucestershire, United Kingdom

[#]Decker.dinesh@gmail.com

ABSTRACT Gender bias in artificial intelligence (AI) systems, particularly within education and workplace settings, poses serious ethical and operational concerns. These biases often stem from historically skewed datasets and flawed algorithmic logic, which can lead to the reinforcement of existing inequalities and the systematic exclusion of underrepresented groups, especially women. This systematic review analyses peer-reviewed literature from 2010 to 2024, sourced from IEEE Xplore, Google Scholar, PubMed, and SpringerLink. Using targeted keywords such as AI gender bias, algorithmic fairness, and bias mitigation, the review assesses empirical and theoretical studies that examine the causes of gender bias, its manifestations in AI-driven decision-making systems, and proposed strategies for detection and mitigation. Findings reveal that biased training data, algorithm design flaws, and unacknowledged developer assumptions are primary sources of gender discrimination in AI systems. In education, these systems affect grading accuracy and learning outcomes; in workplaces, they influence hiring, evaluations, and promotions. Mitigation approaches can be categorized into three main categories: data-centric (e.g., data augmentation and data balancing), algorithm-centric (e.g., fairness-aware learning and adversarial training), and post-processing techniques (e.g., output calibration). However, each approach faces implementation challenges, including trade-offs between fairness and accuracy, lack of transparency, and the absence of intersectional bias detection. The review concludes that gender fairness in AI requires integrated strategies that combine technical solutions with ethical governance. Ethical AI deployment must be grounded in inclusive data practices, transparent protocols, and interdisciplinary collaboration. Policymakers and organizations must strengthen accountability frameworks, such as the EU AI Act and the U.S. AI Bill of Rights, to ensure that AI technologies support equitable outcomes in education and employment.

INDEX TERMS: Artificial Intelligence, Gender Bias, Algorithmic Fairness, Workplace Discrimination, Bias Mitigation in Education

I. INTRODUCTION

The integration of artificial intelligence (AI) into education and workplace systems has introduced both opportunities for efficiency and risks of perpetuating historical biases. Among these risks, gender bias remains a persistent and deeply rooted concern. AI tools used for student assessment, hiring, promotions, and performance evaluations have demonstrated tendencies to replicate and even intensify preexisting gender inequalities. These outcomes are often traced to biased training datasets, non-transparent algorithms, and the absence of fairness-focused design principles [1], [2].

Despite the growing attention to algorithmic fairness, the literature remains fragmented, with few studies providing an integrated view of how gender bias manifests differently across educational and professional AI applications. This review offers a novel contribution by systematically analyzing peer-reviewed research across both sectors, categorizing bias sources, synthesizing detection and mitigation methods, and evaluating the real-world implications of ethical AI frameworks.

By critically examining empirical and theoretical works published between 2010 and 2024, this review aims to bridge disciplinary gaps, inform future AI design, and support policy interventions. It responds to a crucial research need: to develop unified strategies that address gender bias at multiple levels—data, algorithms, and institutional policy.

AI-driven recruitment systems often reflect historical hiring patterns that favoured men, leading to lower selection rates for equally qualified female candidates [3], [4]. Tools trained on male-dominant datasets have rejected resumes containing gender-coded language such as “women’s chess club” [5].

Facial recognition systems exhibit significant accuracy disparities based on gender. Studies have shown lower recognition rates for female faces, particularly those with darker skin tones, due to biased training datasets [6], [7]. These errors not only affect identity verification but also have profound implications for security and law enforcement.

Educational technologies also demonstrate gender bias, particularly in automated grading and adaptive learning systems. Algorithms trained on biased data reflect gendered performance trends, resulting in skewed outcomes that disadvantage female students [8], [9]. Tutoring platforms may recommend more manageable tasks or offer less feedback to female learners, reinforcing gender-based learning disparities [10].

While some progress has been made through fairness-aware algorithms and explainable AI (XAI), implementation remains limited. Tools like Grad-CAM [11] and model cards [12] improve transparency but are rarely adopted in commercial settings [13]. Additionally, fairness frameworks often overlook intersectional

dimensions such as race, class, and disability, narrowing their real-world effectiveness [14].

This paper contributes to the field in three significant ways:

1. **Cross-sector synthesis:** Unlike prior studies focusing exclusively on either education or employment, this review unifies both domains under a single analytical framework.
2. **Methodological rigour:** The study employs a systematic approach to identify, categorize, and critically evaluate the most influential peer-reviewed research published between 2010 and 2024.
3. **Policy relevance:** The review incorporates a discussion of governance frameworks (e.g., EU AI Act, U.S. AI Bill of Rights), providing actionable insights for the implementation of ethical AI.

II. METHODOLOGY

This study employed a systematic review methodology to evaluate peer-reviewed literature related to gender bias in artificial intelligence (AI) systems within educational and workplace contexts. The review followed structured protocols inspired by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [15] framework to ensure transparency and replicability.

A. Data Sources and Search Strategy

A comprehensive search was conducted using four major academic databases: IEEE Xplore, Google Scholar, PubMed, and SpringerLink. The search covered studies published between January 2010 and March 2024, using combinations of the following keywords:

- AI gender bias
- Bias in AI hiring
- Algorithmic fairness in education
- Gender discrimination in AI
- Bias mitigation in machine learning

B. Inclusion and Exclusion Criteria

Inclusion Criteria:

- Peer-reviewed journal articles or conference papers.
- Published between 2010 and 2024.
- Focused on AI applications in education or workplace settings.
- Discussed gender bias detection, impact, or mitigation.
- Provided either empirical findings or theoretical frameworks.

Exclusion Criteria:

- Non-peer-reviewed sources (e.g., blogs).
- Studies unrelated to gender (e.g., focusing only on racial bias).
- Technical papers without social or ethical context.
- Non-English publications.

C. Study Selection and Screening

A PRISMA-style flow diagram [15] summarizing the selection process is provided in Figure 1.

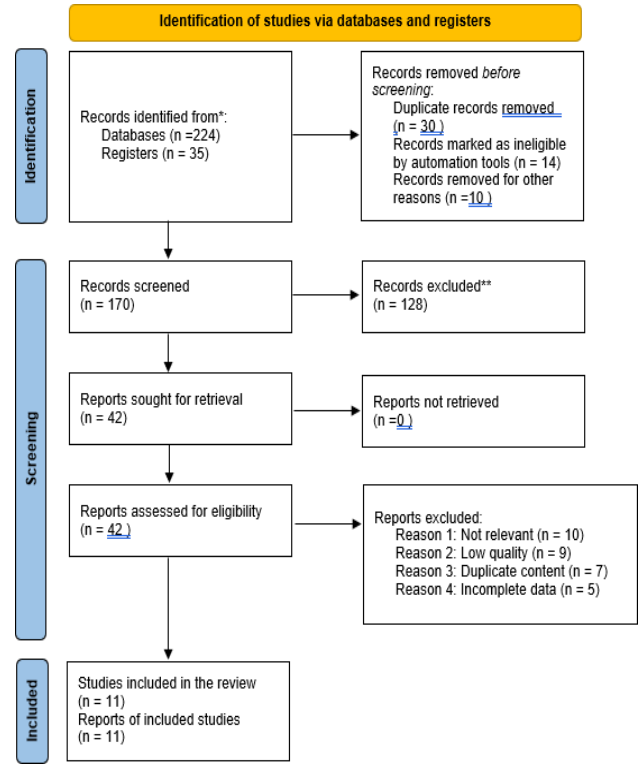


Figure 1: PRISMA 2020 flow diagram outlining the study selection process.

D. Evaluation Framework

To ensure systematic assessment, each selected study was evaluated based on:

- **Contextual domain:** Education or workplace.
- **Bias category:** Data-level, algorithm-level, or outcome-level bias.
- **Mitigation strategies:** Data-centric, algorithm-centric, or post-processing methods.
- **Type of contribution:** Empirical (e.g., experiments, case studies) or theoretical (e.g., frameworks, policy analysis).

The authors also recorded whether studies addressed intersectional bias, discussed ethical implications, and referenced existing governance policies such as the EU AI Act or the U.S. AI Bill of Rights.

III. LITERATURE REVIEW

A. AI's Role in Perpetuating Gender Bias

Culture-biased training data generates artificial intelligence systems that replicate and amplify such social biases, as reported by Ntoutsis [16], Kchling [19], and Slimi [8]. AI systems that use machine learning algorithms draw knowledge from extensive datasets but reproduce and magnify biases within them through their production outputs [1]. AI recruitment tools that train using historically biased information will disadvantage the selection of female candidates [3],[4]. Education technologies, including admission and grading systems, operate with potential gender bias due to data presentation of current academic performance gaps between genders [8], [9]. The data origin finally leads to universal bias problems affecting all educational and work-related areas [18]. According to Shrestha and Das, the design workflow for

algorithms produces systematic biases that are incorporated into the final products [2].

The application of AI in facial recognition systems produces discriminatory results that affect different genders, according to [6]. Using datasets that primarily feature male faces results in systems producing reduced accuracy for female face identification, which can lead to analysis errors [7]. The unreliable nature of these systems may have significant societal consequences in security fields and law enforcement areas, which can exacerbate discrimination [12]. Such biased systems necessitate immediate attention regarding their legal and ethical implications, according to Ntoutsis [16].

B. Methods for Detecting and Mitigating Gender Bias in AI Systems

Multiple scholarly works are devoted to AI gender bias detection and mitigation methods, according to Shrestha [2], Liu [7], and Holstein [13]. Different detection approaches and applications manifest into distinct strategies for these methods. Research shows that analyzing training data for gender biases constitutes a standard method [20],[21]. The evaluation process includes recognising and fixing data distribution faults that prevent correct population representation. Data augmentation represents an explored technique that increases underrepresented population groups through artificial methods [21]. The algorithms can be modified through specific adjustments that reduce their sensitivity to gender-related features [7]. The development of algorithms should focus on two approaches: adding fairness constraints during learning and improving capabilities to resist biases in data.

It is essential to develop explainable AI (XAI) methods to understand how AI models perform processes and locate potential biases, according to Asatiani [22] and Hassija [23]. AI transparency becomes possible through XAI methods, which enable researchers and practitioners to understand the factors that affect model predictions and identify the origins of bias. Model prediction explanations derived from Grad-CAM [11] generate images that help users identify biases within model representations. Model cards introduced by Mitchell [12] help organisations maintain transparency through the documentation of model performance data, which includes results from different gender groups, making it easier to detect biases. A considerable barrier exists because commercial product development teams face limitations in the proposed solutions presented in fair ML research literature [13].

AI algorithms now analyze educational content so teachers can identify gender misconceptions to create balanced learning spaces between the genders [2]. Artificial intelligence develops tools that deliver customised assessments to learners to achieve gender-balanced educational achievement [10]. Data privacy concerns related to algorithmic bias should be diligently addressed when developing these systems [24]. AI education necessitates a human-centred approach to ensure the

development and implementation of technology that fosters fairness and equity [18].

C. AI in Gender Bias Detection and Mitigation in Workplaces

The workplace utilises AI technology to streamline recruitment processes, evaluate performance, and make promotion decisions. AI imposes gender biases on these decisions unless proper management is implemented, according to Hunkenschroer [3] and Ferrer [4]. AI recruiting tools that receive inputs from biased data systems will reject eligible female candidates, according to Shrestha [2] and Ferrer [4]. AI systems that evaluate performance can replicate existing gender biases in performance measures, leading to discriminatory evaluation assessments [9]. AI systems possess the capability to find gender bias issues at work sites and establish methods to reduce the impact of bias. AI algorithms generate insights about gender-biased wordings in job descriptions, which enables businesses to enhance their recruiting materials, according to Shrestha [2]. AI monitoring tools track workplace interactions to identify signs of bias, enabling organisations to develop better workplace equity practices [25]. Organizations must handle AI workplace deployment through attention to ethical issues that combine data privacy risks with bias concerns found in algorithmic systems [26].

D. Research Gap

While there is growing scholarly attention to the ethical and technical aspects of gender bias in AI systems, existing reviews often focus narrowly on either algorithmic fairness in general or gender discrimination in isolated contexts such as hiring or facial recognition. These studies typically overlook the combined impact of gender bias across both education and workplace environments, which are increasingly interconnected through AI-driven decision-making tools.

Furthermore, many prior reviews emphasize detection and mitigation strategies but fall short of integrating policy frameworks and ethical governance models into their analysis. The lack of attention to intersectional bias, where gender bias overlaps with other dimensions such as race, socioeconomic status, or disability, also leaves critical gaps in understanding how AI systems affect different groups simultaneously.

Our review addresses these deficiencies by:

- Synthesising literature from both educational and employment contexts within a single framework.
- Categorizing sources, impacts, and mitigation techniques of gender bias in a structured, comparative format.
- Highlighting the role of recent policy developments (e.g., EU AI Act, U.S. AI Bill of Rights) in shaping ethical responses to gender bias in AI.
- Calling for intersectional approaches to bias detection and mitigation.

By bridging disciplinary silos and connecting technical, ethical, and institutional perspectives, this review offers a more comprehensive understanding of gender bias in AI an essential step toward the equitable and accountable deployment of AI in real-world settings.

IV. KEY FINDINGS

This section synthesises findings from 11 representative studies selected for their detailed insights into bias types, mitigation strategies, intersectionality considerations, and policy frameworks relevant to AI applications in education and workplace settings.

A. Evaluation Dimensions and Framework

Each study was evaluated across five key dimensions:

- **Domain:** The primary focus area Education, Workplace, or Both.
- **Bias Category:** The level at which bias manifests Data, Algorithmic, or Outcome.
- **Mitigation Strategy:** The corrective or preventative approach Data-centric, Algorithm-centric, Post-processing, or Policy-based.
- **Intersectionality:** Whether intersecting axes of discrimination (e.g., gender + race) were considered.
- **Policy Framework:** Whether the study aligned with or proposed formal governance strategies.

This evaluation matrix facilitated consistent classification across studies and provided a foundation for comparative analysis.

B. Domain and Contribution Type Distribution

Among the 11 analyzed studies:

- 6 studies focused on workplace bias, particularly algorithmic discrimination in recruitment systems, Ex:[5], [4]
- 3 studies addressed educational bias, including grading algorithms and adaptive systems, Ex:[7], [10].
- 2 studies spanned both domains, analyzing systemic and multi-level biases, Ex:[2]

These studies include both empirical (e.g., dataset evaluations, model testing) and theoretical contributions (e.g., policy reviews, fairness frameworks).

C. Bias Categories and Mitigation Strategies

Biases were categorised and addressed as follows:

Bias Type:

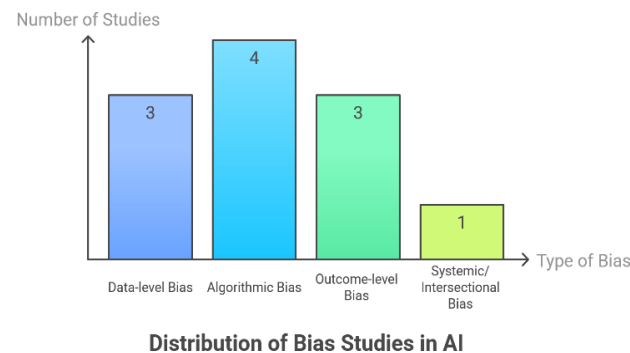


Figure 2: Distribution of bias types identified in the reviewed studies: algorithmic bias (n = 4), data-level bias (n = 3), outcome-level bias (n = 3), and systemic/intersectional bias (n = 1).

Mitigation Strategies:

Some studies adopted hybrid approaches, addressing both technical and governance-level interventions.

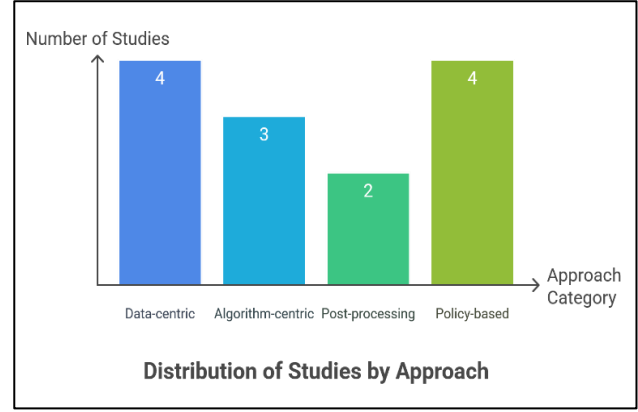


Figure 3: Distribution of included studies by mitigation approach category: data-centric (n = 4), algorithm-centric (n = 3), post-processing (n = 2), and policy-based (n = 4).

D. Study Quality Assessment

Assessment was based on scope, methodological transparency, and practical relevance:

Table 1: Study Quality Assessment Based on Methodological Rigour and Scope

Quality Tier	No. of Studies	Description
High	4	Multi-method, large datasets, applied policy frameworks
Medium	5	Methodologically sound but context-limited
Low	2	Conceptual only or lacked empirical grounding

E. Sources of Gender Bias in AI

Multiple interrelated factors contribute to gender bias within artificial intelligence systems, amplifying each other's impact. AI training data contains systematic gender discrimination because it draws information from historical databases that replicate social imbalances between men and women. AI recruitment tools that learn from historical hiring data that disproportionately favoured men will continue the biased behaviour [16], [12]. The accuracy of facial recognition systems drops among female identification when their training datasets disproportionately favour males, according to Mitchell [12]. When trained on biased text, corpus language models tend to adopt gender stereotypes reflected during operation [27].

The basic design of AI systems prioritises operational efficiency over fair treatment. Design solutions developed during feature selection, alongside optimisation criteria, risk producing discriminatory evaluation results across hiring assignments, assessments, and promotion decisions [4], [17].

AI development processes heavily depend on the biases that developers insert throughout the construction phase. Model deployment techniques, training, and testing phases

depend on developers who might not be aware that their implicit biases affect the process. Female and male developers experience stereotyped outcomes because projects often lack diverse teams and utilise biased-unaware programs, as identified by O'Connor [1] and Shrestha [2]. The solution to these difficulties needs intentional action to create equitable artificial intelligence systems, which must incorporate diverse representation and transparent systems and procedures to fight bias.

F. Impact of Gender Bias in AI on Education and Workplaces

AI systems across educational settings and workplaces maintain discriminatory behaviours because of gender bias; thus, they reinforce opportunity inequality.

AI tools designed for educational evaluation and customisation reinforce gender prejudice, so students receive discriminatory feedback and encounter educational environments that systematically favour males. Educational datasets with prejudicial bias cause tutoring systems to provide inadequate support to female students, negatively impacting their educational development [28]. According to Popenici [29], automated grading systems and language models benefit primarily male-dominated academic institutions by favouring female students.

Implementing biased AI systems within workplaces can lead to unfair discrimination throughout the hiring process, evaluation methods, and promotional advancement criteria. Hiring tools that utilise artificial intelligence and train with data, often showing a male predominance, may prevent female candidates from progressing or rank them lower [17]. Artificial intelligence systems that use automated performance evaluations tend to provide superior evaluation scores to male workers, which negatively affects their compensation and professional growth [1]. Gender-specific biases within leadership decisions actively promote inequalities between men and women according to workplace authority and salary distribution, and reduce the opportunities for women's career growth.

Future improvement demands precise methods of operation and frameworks that account for fairness, as well as various representations during AI development, to bring equitable opportunities in educational institutions and professional careers.

G. Mitigation Strategies for Gender Bias in AI

Research and development initiatives have identified three primary routes for mitigating gender biases arising from Artificial Intelligence systems, encompassing data-centric, algorithm-centric, and post-processing strategies. These mitigation approaches work at various points throughout AI development to establish fairness and eliminate bias in decisions made by Artificial Intelligence systems.

Training datasets must be adequately balanced and contain diverse datasets to achieve unbiased AI outputs according to data-centric approaches. Gender diversity in automated systems benefits from data augmentation techniques that identify and eliminate biases in their source [4]. Preparing datasets with proper demographic representation ensures a

reduction in bias in AI systems that have not yet been disseminated. The quality investments and representative efforts to train data enable AI systems to understand equitable decision-making patterns during learning processes.

AI models become fairer when algorithm-centric approaches add fairness-aware decision-making functions during modifications of AI models. AI models should integrate gender neutrality into their systems by creating models that actively recognise unfairness and employ adversarial training to remove biased pattern outputs [30]. Fairness constraints integrated into the training process enable AI systems to evaluate equitable outcomes during decision-making intentionally [2]. The modifications enable fairer algorithmic processing, reducing AI model tendencies to perpetuate existing gender disparities.

The application of bias-aware modifications occurs after artificial intelligence systems create their prediction results through post-processing methods. The process of calibrating AI-generated outcomes provides corrections against biased hiring and grading practices, and fair ranking systems block AI from showing a preference for male candidates [4], [17]. The effectiveness of post-processing methods at minimizing immediate biases does not solve underlying biases found in training data and algorithms. The long-term achievement of fairness in AI systems heavily depends on receiving immediate attention from data-centric and algorithm-centric solutions systems. AI developers should implement various mitigation approaches to develop AI-driven decision systems that support fair and unbiased practices.

H. Challenges in Implementing Bias Mitigation Strategies

Multiple real-world obstacles prevent the deployment of available bias mitigation tools during decision-making processes that rely on AI systems. Enhancing fairness often means that AI will have reduced efficiency and decreased accuracy. Academic and professional design choices need ethical standards to keep AI performance and fairness at acceptable levels [30]. Many AI systems face ethical problems and transparency issues because they lack clear procedures for bias detection, fairness assessment, and accountability monitoring. The lack of sufficient AI governance frameworks necessitates those policymakers develop new regulations to maintain transparency and explainability, thereby establishing trust in AI-based decision-making [1]. Most AI fairness techniques only evaluate gender-based biases, yet they fail to address combined biases, which include those related to race, ethnicity, and socioeconomic status. Stand-alone AI systems require programming that enables them to identify multiple layers of discrimination factors and prevent unfair treatment of different population groups [14].

AI tools in education show promise for individualised learning and better results, but biased systems perpetuate gender-based prejudices, which result in unequal instructional approaches [28], [29]. Feminine students face disadvantages when taking tests through AI-powered tutoring platforms and automated grading tools, as these systems often support writing formats and communication patterns that are not inclusive of women [17]. Reducing

risks in AI systems demands transparency, accountability features, and fairness design principles. Students and instructors should actively collaborate on AI system development so that all learners experience unbiased and equal educational settings [1], [29].

The use of AI systems to recruit personnel and assess employee performance during promotions often exacerbates gender discrimination unless AI frameworks are designed explicitly to prevent it. The selection tool, which utilises AI-powered analysis of biased data points, disproportionately screens out female candidates. At the same time, performance evaluation algorithms with embedded gender stereotype logic show a preference toward male employees, according to Raghavan [17] and Booth [31]. Fairness-aware algorithms, in combination with representative datasets and adequate evaluation techniques, help detect and reduce prejudice in technology systems [4]. The deployment of AI technologies requires the promotion of diversity and inclusive policies to ensure equitable job satisfaction and workplace fairness among workers [17].

Different strategies to reduce gender bias in AI systems include programmatic solutions that focus on various stages, from development to execution. The validity of training data must remain balanced and diverse, as data-centric approaches aim to eliminate bias. Combining data augmentation with bias audits and representative dataset curation techniques addresses biases at their root to prevent inherited societal inequalities in AI systems [4]. The modification of AI models through fairness-aware algorithms and adversarial training techniques with embedded fairness constraints during model training constitutes algorithm-centric approaches, according to Meade [30] and Booth [31]. After AI predicts results, post-processing methods apply corrections to the system output for hiring processes, grading, and system ranking functions to reduce biases. AI avoids gender bias discrimination by implementing prediction calibration techniques and fair ranking methods [4], [17]. Even though these bias reduction methods yield instant results, they fail to address fundamental systematic bias; therefore, lasting solutions must begin with data collection and extend to algorithm design.

Amazon's AI recruitment system demonstrated a significant trade-off, as it discriminated against female candidates while favouring male candidates. 2014 marked Amazon's creation of AI recruitment technology that scanned candidates' qualifications and positioned them through resume analytics. The recruitment system acquired knowledge from historical employment data, which predominantly contained male applicants, as the tech field was predominantly male-dominated during that period. The AI system decided to give lower rankings to resumes containing terms related to women, such as "women's" (e.g., "women's chess club"), while prioritising male-heavy experiences and occupational language [5].

When changing its programming, the biased algorithm forced Amazon to struggle between operational efficiency and fairness goals. The system training to reach fairness

goals resulted in diminished performance from the AI model. The AI tool did not launch after Amazon phased it out in 2018 because the company found it too burdensome to connect accurate hiring decisions and unbiased operations [5]. Balancing the performance quality of AI systems with solution-based fairness remains a significant challenge. At the same time, tech teams handle deep-seated biases in their training data.

The recognition of steady fairness audits proves that AI modelling depends on human oversight and regulatory oversight to prevent biased outcomes while protecting operational efficiency. Working seriously with transparency and auditing operations on training data types enables bias prevention without compromising operational AI efficiency [32].

AI will reach its maximum potential in education and employment through continuous efforts to solve gender bias concerns. Creating diverse teams for software development and implementing transparency systems with fairness-conscious AI solutions form necessary elements for making fair AI applications. Ethical AI governance, which uses diverse data coupled with thorough bias-minimisation approaches, makes AI an instrument that builds more just and inclusive digital settings.

1. Ethical Challenges and Policy Considerations in AI Bias Mitigation

The implementation of ethical guidelines, combined with disclosure measures and regulatory approaches, protects against gender bias while preventing further types of discrimination in AI-generated decision-making processes. Research on AI fairness has progressed, although fundamental governance challenges persist due to AI systems' significant influence over the educational and employment sectors. Government bodies, professional groups, and private organizations have the key duty to create standards for AI fairness.

The global regulations for AI fairness continue to evolve through new legislative frameworks that focus on bias detection, alongside requirements for transparency, accountability systems, and ethical AI governance. As one of its most advanced projects, the European Union implemented the AI Act (2021), which categorises AI systems by risk levels and then mandates detailed bias evaluations and complete transparency for all high-risk AI systems operating in employment, educational assessment and law enforcement tasks [33]. All AI deployments handling these domains must comply with fairness and non-discrimination standards through conformity assessments. The U.S. Blueprint for an AI Bill of Rights (2022) establishes parameters to protect AI safety and promote fairness and accountability through demands for bias examination, human supervision systems, and protection against discrimination in hiring, education, and financial domain applications [34]. The framework serves as a recommendation for AI developers and policymakers seeking to establish fairness protections in AI regulation.

Public authorities, private industry, and scientific research institutions are working together to mitigate AI bias and

develop inclusive AI governance frameworks. Regulatory bodies require authority to enforce AI impact assessments, conduct bias detection audits, and maintain transparency standards to ensure compliance with fairness protocols. The EU AI Act requires companies to demonstrate the safety of their high-risk AI systems through "conformity assessments," which involve showing that they do not cause excessive harm to specific demographic groups [33]. Leading corporations such as Google, Microsoft, and IBM have developed AI fairness frameworks that include routine hiring tool assessments, employ bias identification mechanisms, and distribute their AI technology using fair models [32]. Such programs exemplify how private companies can implement initiatives to reinforce government policies that mitigate the impact of AI bias. Standardized bias detection approaches and mitigation frameworks require a joint effort between AI researchers to work with ethicists who connect with legal experts and policymakers in developing these systems. Developing artificial intelligence through multi-stakeholder partnerships ensures that technical developments align with ethical framework standards and legal and societal fairness principles [35].

AI systems must maintain ethical integrity through principled AI designs, fairness-aware programming methods, and inclusive data management practices to promote fairness and accountability. As Floridi [35] pointed out, bias audits and transparency evaluations, along with algorithmic explainability tests, should become regular procedures for maintaining bias-free and interpretable decision-making processes. Individuals in street enterprises can utilise "algorithmic fairness scorecards" to evaluate AI performance data across various population categories, enabling the identification of bias origins. AI developers need to employ training datasets incorporating diverse representations to reduce bias patterns in the data. Data augmentation combined with fairness-aware sampling and intersectional bias analysis allows organisations to minimise discrimination during AI decision-making processes [32]. Administrative supervision systems must operate within AI-powered recruitment systems, while educational and vital decision-making fields require human oversight to eliminate automatic unfair treatment. Establishing ethics review boards, AI transparency reporting requirements, and fairness auditing standards allow organisations to become responsible when they generate biased AI outcomes [34].

The regulation of gender bias in Artificial Intelligence requires multiple approaches that combine standardized policies with business accountability and diverse partnerships among professionals. Nationals should create AI impact assessment requirements through bias auditing legislation that mandates corporations to establish independent methods for ensuring fairness and providing explanation tools in their systems. AI governance systems require ethical processes combined with policy tools for transparent data and inclusive practices, as they prevent the retention of social bias in AI systems and uphold justice. Organisations can build trust in AI technology through the combination of policy-oriented supervision and ethical AI

governance standards, creating education systems and workplace environments that promote greater fairness.

J. Summary of Key Findings

- Workplace studies revealed predominant data and algorithmic biases affecting recruitment outcomes, e.g., [5], [17].
- Education studies highlighted challenges in algorithm fairness and outcome disparities, e.g., [7], [8].
- Policy-integrated research, e.g., [12], [16] showcased frameworks such as model cards and fairness audits.
- Intersectionality was explicitly addressed in only a few studies, pointing to a need for deeper multidimensional analyses.
- While mitigation strategies are maturing, the field still lacks longitudinal evaluations of their effectiveness and scalability.

V. DISCUSSION

This review confirms that gender bias remains a persistent challenge in AI applications across both educational and workplace contexts. While the reviewed literature reflects growing awareness and sophistication in identifying and addressing bias, the effectiveness of proposed mitigation strategies varies significantly.

A. Critical Reflection on Mitigation Strategies

Data-centric approaches, such as data augmentation and rebalancing, are widely used (e.g., [4], [16]), but they rely heavily on the assumption that bias is primarily rooted in the dataset. This overlooks structural and historical inequalities that shape the data in the first place. Additionally, these methods can unintentionally oversample minority representations, leading to distorted distributions or performance trade-offs.

Algorithm-centric methods, such as fairness-aware training and adversarial debiasing (e.g., [7], [3]), show promise in improving model behaviour during training. However, their implementation often requires advanced technical expertise and computational resources, which are not equally available across all organizations. Moreover, many of these models operate as "black boxes," reducing interpretability and user trust [13], [23].

Post-processing techniques, such as output calibration and ranking correction (e.g., [17]), are relatively more straightforward to implement but are reactive rather than preventive. They treat the symptoms of bias after decisions are made rather than addressing underlying causes, and their effectiveness is typically limited to the specific application without generalizability.

Policy-driven strategies such as model documentation [12] and fairness audits [32] are essential for accountability. However, uptake is inconsistent across sectors, and few policies are enforceable. Intersectional bias—addressed by only a minority of studies (e.g., [14])—remains a critical gap, especially when AI systems interact with overlapping axes of discrimination such as race, class, or disability.

Table 2: Summary of Mitigation Strategies with Examples, Advantages, and Limitations

Mitigation Strategy	Examples	Advantages	Limitations
Data-Centric	Data audits, rebalancing, augmentation [4], [16]	Addresses bias at the source	May reinforce structural inequalities; data availability
Algorithm-Centric	Fairness-aware training, adversarial debiasing [31], [2]	Tackles bias during model training	Requires technical expertise; interpretability issues
Post-Processing	Score calibration, fair ranking [17]	Easy to implement post hoc	Reactive, not preventive; limited scope
Policy-Based	Model cards, ethics audits, transparency tools [12], [14]	Enables accountability and governance	Enforcement is weak; adoption inconsistent
Explainability (XAI)	SHAP, LIME, Grad-CAM [18]	Enhances transparency and trust	Often only diagnostic, not corrective
Intersectional Analysis	Multi-dimensional bias evaluation [14], [8]	Reveals layered inequalities	Rarely applied; complex to operationalize

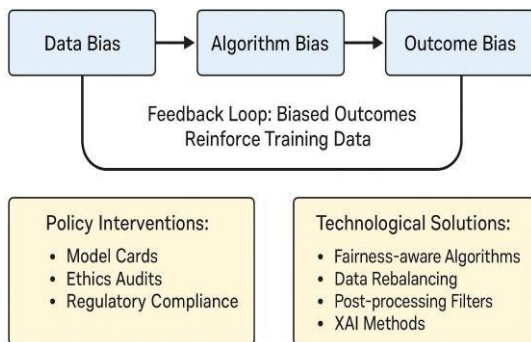


Figure 4: Conceptual framework illustrating the cycle of bias in AI systems. Data bias propagates into algorithmic bias, resulting in outcome bias. A feedback loop reinforces training data with biased outcomes. Interventions are categorised into policy-based (e.g., model cards, ethics audits, regulation) and technological solutions (e.g., fairness-aware algorithms, data rebalancing, XAI).

VI. FUTURE RESEARCH DIRECTIONS

Research on bias prevention for AI should focus on three fundamental areas: intersectional fairness, ethical AI development, and real-world impact assessment. For spotting AI biases, research needs to establish gender bias analysis concerning other forms of discrimination, such as race, social position, and disabilities. AI systems must train their ability to recognise and resolve several biases in parallel operations to generate complete fairness results. A systematic analysis in AI development that supports multiple identities helps mitigate simultaneous discrimination issues that often affect minorities.

Future ethical frameworks designed for AI require development to produce enforceable rules for gender fairness throughout the AI development process. Proactive fair AI programs require mechanisms to combine bias identifications with ethical protocols, establishing transparency procedures for maintaining fairness consistency. Trust in AI systems influencing hiring operations, grading, and promotion algorithms will be established by aligning explainability with accountability standards. Software developers creating AI systems should adopt technologies from explainable AI (XAI) that enable organisations, along with users, to gain insight into automated decisions and evaluate the fairness of their results. Situations involving critical decisions necessitate heightened importance because biased AI-driven decisions lead to severe educational and professional results for individuals.

The scientific research about gender bias in AI continues to expand, yet important information gaps persist. To better understand the permanent societal transformations from biased AI systems and the performance of different bias reduction methods across multiple fields, researchers need to conduct additional studies [4], [30]. The ethical consequences of AI in education and the workplace require further investigation, as transparency, fairness, and accountability become significant concerns, according to [1]. Exploring bias requires a deeper study of intersectionality because it describes how gender bias operates alongside social categories like race, ethnicity, and socioeconomic status per Guo [14]. Research initiatives must advance systematic approaches to identify bias in AI systems while exploring moral practices in AI development and implementing practical deployment methods [4].

Implementing AI bias mitigation strategies requires evaluation through time-dependent studies, as schools and workplaces necessitate assessments of their enduring effects. Short-term experimental settings characterise most current AI fairness studies because they fail to show long-term performance outcomes for fairness interventions. Real-world controlled assessments across substantial application domains will generate essential proof about bias reduction methods, enabling policymakers, organization leaders, and AI developers to establish their best practices.

Progress in resolving AI bias is priceless and depends on essential cooperation between AI scientists, social experts, ethicists, and government officials [1], [4]. Multiple

disciplines must collaborate to ensure the fairness, transparency, and ethical compliance of AI decision-making processes that address complex biases. Inclusive AI governance mechanisms should be established to create rules that ensure AI algorithms adhere to fairness principles by promoting inclusive practices in educational and work environments. AI fairness research that combines collaborative methods and broad institutional approaches will successfully reduce gender bias as it develops ethically responsible AI technology.

VII. CONCLUSION

This systematic review analyzed 11 peer-reviewed studies spanning 2010–2024 to examine how gender bias manifests in AI systems and how such bias is detected and mitigated. The review encompassed applications in both education and the workplace, offering a comprehensive perspective across domains where AI-driven decisions can significantly impact individual opportunity and equity.

The findings show that:

- Gender bias originates from biased training data, flawed algorithms, and a lack of ethical oversight.
- Mitigation strategies fall into three main categories: data-centric, algorithm-centric, and post-processing, with emerging support for policy-level governance.
- Many reviewed studies highlight the trade-off between fairness and performance, and a lack of intersectional bias detection persists.
- Long-term, real-world evaluations of fairness interventions are notably absent, limiting the field's ability to gauge sustainable impact.

The most substantial contributions come from studies that integrate technical and ethical perspectives, such as those by Shrestha and Das [2], Mitchell et al. [12], and O'Connor and Liu [1]. These works advocate not only for improved models but also for structural changes in how AI is regulated, developed, and audited.

To move toward equitable AI systems, future work must:

- Invest in explainable AI (XAI) tools that make fairness visible and actionable.
- Mandate policy compliance mechanisms, such as those introduced in the EU AI Act and the U.S. AI Bill of Rights.
- Expand the lens of analysis to include intersectionality, ensuring that AI systems do not disproportionately harm already marginalized communities.

Ultimately, fair AI is not only a technical challenge but a societal one requiring collaboration between engineers, policymakers, educators, ethicists, and affected communities.

REFERENCES

- [1] S. O'Connor and H. K. Liu, "Gender bias perpetuation and mitigation in AI technologies: Challenges and opportunities," **AI & Society**, vol. 38, pp. 917–933, 2023, doi: 10.1007/s00146-023-01675-4.
- [2] S. Shrestha and S. Das, "Exploring gender biases in ML and AI academic research through systematic literature review," **Frontiers in Artificial Intelligence**, vol. 5, 2022, doi: 10.3389/frai.2022.976838.
- [3] A. L. Hunkenschroer and C. Luetge, "Ethics of AI-enabled recruiting and selection: A review and research agenda," **Journal of Business Ethics**, vol. 182, pp. 243–261, 2022, doi: 10.1007/s10551-022-05049-6.
- [4] X. Ferrer, T. V. Nuenen, J. M. Such, M. Cot, and N. Criado, "Bias and discrimination in AI: A cross-disciplinary perspective," **IEEE Technology and Society Magazine**, vol. 40, no. 1, pp. 72–80, 2021, doi: 10.1109/MTS.2021.3056293.
- [5] J. Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," **Reuters**, Oct. 10, 2018. [Online]. Available: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
- [6] P. Terh rst *et al*., "A comprehensive study on face recognition biases beyond demographics," **IEEE Transactions on Technology and Society**, vol. 2, no. 4, pp. 199–212, 2021, doi: 10.1109/TTS.2021.3111823.
- [7] H. Liu, J. Dacon, W. Fan, H. Liu, Z. Liu, and J. Tang, "Does gender matter? Towards fairness in dialogue systems," in *Proc. Int. Conf. Computational Linguistics (COLING)*, Barcelona, Spain, Dec. 2020, pp. 4405–4415. doi: 10.18653/v1/2020.coling-main.390.
- [8] Z. Slimi and B. Villarejo-Carballido, "Navigating the ethical challenges of artificial intelligence in higher education: An analysis of seven global AI ethics policies," *TEM Journal*, vol. 12, no. 2, pp. 548–554, 2023. doi: 10.18421/TEM122-02.
- [9] L. Cheng, K. R. Varshney, and H. Liu, "Socially responsible AI algorithms: Issues, purposes, and challenges," *Journal of Artificial Intelligence Research*, vol. 71, pp. 1089–1121, 2021. doi: 10.1613/jair.1.12814.
- [10] F. Kamalov, D. S. Calonge, and I. Gurrib, "New era of artificial intelligence in education: Towards a sustainable multifaceted revolution," *Sustainability*, vol. 15, no. 16, pp. 12451, 2023. doi: 10.3390/su151612451.
- [11] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74.
- [12] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, et al., "Model cards for model reporting," in *Proc. Conf. Fairness, Accountability, and Transparency (FAT)*, Atlanta, GA, USA, Jan. 2019. doi: 10.1145/3287560.3287596.
- [13] K. Holstein, J. W. Vaughan, H. Daum  III, M. Dudik, and H. Wallach, "Improving fairness in machine learning systems: What do industry practitioners need?" in *Proc. 2019 CHI Conf. Human Factors*

- Comput. Syst.*, Glasgow, Scotland, May 2019, pp. 1–16. doi: 10.1145/3290605.3300830.
- [14] S. Guo, J. Wang, L. Lin, and R. Chen, "The impact of cognitive biases on decision-making processes in high-stress environments," *Journal of Cognitive Psychology*, vol. 33, no. 5, pp. 567–580, 2021.
- [15] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, no. n71, pp. 1–9, 2021. doi: 10.1136/bmj.n71.
- [16] E. Ntoutsis, P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdl, M. Vidal, et al., "Bias in data-driven artificial intelligence systems: An introductory survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, pp. e1356, 2020. doi: 10.1002/widm.1356.
- [17] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring: Evaluating claims and practices," in *Proc. Conf. Fairness, Accountability, and Transparency (FAT)*, Barcelona, Spain, Jan. 2020, pp. 469–481. doi: 10.1145/3351095.3372873.
- [18] S. J. Yang, H. Ogata, T. Matsui, and N. Chen, "Human-centered artificial intelligence in education: Seeing the invisible through the visible," *Cognitive and Affective Computing*, vol. 2, no. 1, pp. 1–14, 2021. doi: 10.1016/j.caeai.2021.100008.
- [19] A. Küchling and M. C. Wehner, "Discriminated by an algorithm: A systematic review of discrimination and fairness in algorithmic decision-making in HR recruitment and development," *AI and Ethics*, vol. 1, pp. 1–17, 2020. doi: 10.1007/s40685-020-00134-w.
- [20] A. Thieme, D. Belgrave, and G. Doherty, "Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems," *ACM Trans. Comput.-Hum. Interact.*, vol. 27, no. 5, pp. 1–53, 2020. doi: 10.1145/3398069.
- [21] A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, "Data and its (dis)contents: A survey of dataset development and use in machine learning research," *Patterns*, vol. 2, no. 11, pp. 100336, 2021. doi: 10.1016/j.patter.2021.100336.
- [22] A. Asatiani, P. Malo, P. R. Nagbl, E. Penttinen, T. Rinta-Kahila, and A. Salovaara, "Challenges of explaining the behavior of black-box AI systems," *Journal of Management Science and Quantitative Economics*, vol. 6, no. 1, pp. 1–23, 2020. doi: 10.17705/2msqe.00037.
- [23] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, et al., "Interpreting black-box models: A review on explainable artificial intelligence," *Cognitive Computation*, 2023. doi: 10.1007/s12559-023-10179-8.
- [24] A. Nguyen, H. N. Ngo, Y. Hong, B. Dang, and B. T. Nguyen, "Ethical principles for artificial intelligence in education," *Education and Information Technologies*, vol. 27, pp. 13573–13593, 2022. doi: 10.1007/s10639-022-11316-w.
- [25] M. Mirbabaie, F. Brünker, N. Frick, and S. Stieglitz, "The rise of artificial intelligence: Understanding the AI identity threat at the workplace," *Electronic Markets*, vol. 31, pp. 895–913, 2021. doi: 10.1007/s12525-021-00496-x.
- [26] White House Office of Science and Technology Policy (OSTP), "Blueprint for an AI Bill of Rights: Making automated systems work for the American people," Washington, DC, USA, 2022. [Online]. Available: <https://www.whitehouse.gov/ostp/ai-bill-of-rights>.
- [27] P. Budhwar, S. Chowdhury, G. Wood, H. Aguinis, G. J. Bamber, J. R. Beltran, et al., "Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT," *Human Resource Management Journal*, vol. 34, no. 1, 2023. doi: 10.1111/1748-8583.12524.
- [28] A. Caliskan, P. A. Pimparkar, T. Charlesworth, R. Wolfe, and M. R. Banaji, "Gender bias in word embeddings: A comprehensive analysis of frequency, syntax, and semantics," in *Proc. 2022 AAAI/ACM Conf. AI, Ethics, and Society (AIES '22)*, Oxford, UK, 2022, pp. 172–182.
- [29] M. Roshanaei, "Cybersecurity preparedness of critical infrastructure: A national review," *Journal of Critical Infrastructure Policy*, vol. 4, no. 1, Article 4, 2023.
- [30] S. Popenici, "The critique of AI as a foundation for judicious use in higher education," *Journal of Applied Learning & Teaching*, vol. 6, no. 2, pp. 378–384, 2023.
- [31] N. Meade, E. Poole-Dayana, and S. Reddy, "An empirical survey of the effectiveness of debiasing techniques for pre-trained language models," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, Dublin, Ireland, May 2022, pp. 1878–1898.
- [32] B. Booth, L. Hickman, S. K. Subburaj, and S. K. D'Mello, "Bias and fairness in multimodal machine learning: A case study of automated video interviews," in *Proc. 2021 ACM Conf. Fairness, Accountability, and Transparency (FACCT '21)*, Virtual Event, Mar. 2021, pp. 279–289.
- [33] I. D. Raji, A. Smart, R. N. White, M. Mitchell, T. Gebru, B. Hutchinson, J. Smith-Loud, D. Theron, and P. Barnes, "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing," in *Proc. 2020 Conf. Fairness, Accountability, and Transparency (FAT)*, Barcelona, Spain, Jan. 2020, pp. 33–44. doi: 10.1145/3351095.3372873.

- [34] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, and E. Vayena, "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018. doi: 10.1007/s11023-018-9482-5.
- [35] European Commission, "Proposal for a regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act)," European Commission, Brussels, Belgium, 2021. [Online]. Available: <https://digitalstrategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>.

AUTHOR BIOGRAPHIES

Dinesh Deckker

Dinesh Deckker is a postgraduate researcher currently pursuing a PhD in Marketing. He holds a BA (Hons) in Business from Wrexham University, UK; an MBA from the University of Gloucestershire, UK; a BSc (Hons) in Computer Science from IIC University of Technology, Cambodia; and an MSc (Hons) in Computing from Wrexham University. His research interests include Artificial Intelligence, Social Sciences, and Linguistics. **ORCID:** <https://orcid.org/0009-0003-9968-5934>

Subashini Sumanasekara

Subashini Sumanasekara is a postgraduate researcher with a strong interdisciplinary background in computing and education. She holds a BSc (Hons) in Computing from the University of Gloucestershire, UK and MSc (Hons) in Strategic IT Management from the University of Wolverhampton, Cambodia and MA (Hons) in Education from Girne American University, Cyprus. Her research interests include Artificial Intelligence, Social Sciences and Linguistics. **ORCID:** <https://orcid.org/0009-0007-3495-7774>

Augmented Reality (AR) and Virtual Reality (VR) in Education: A Comprehensive Review

KMHL Konara^{1#}, GK Dilani¹, TMHC Peiris¹, RMR Dileka¹, TP Rathnayaka¹, WMCJT Kithulwatta¹, RMD Jayathilake², YNS Wijewardana³, HMCC Somarathna⁴, RMKT Rathnayake⁵,

¹ Department of Information and Communication Technology, Faculty of Technological Studies, Uva Wellassa University of Sri Lanka, Badulla, Sri Lanka

² Faculty of Indigenous Medicine, University of Colombo, Sri Lanka

³ Department of Engineering Technology, Faculty of Technological Studies, Uva Wellassa University of Sri Lanka, Badulla, Sri Lanka

⁴ Department of Civil Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, 10115, Sri Lanka

⁵ Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

#ict20090@std.uwu.ac.lk

ABSTRACT Now the world is fully moving towards a digitalized environment in all kinds of disciplines in education, agriculture, the banking industry, transportation, healthcare, etc., with the most prominent and trending topics. Augmented Reality (AR) and Virtual Reality (VR) technologies are contemporary tools that are gradually changing learning processes among plenty of newly arrived technologies. These tools enable educators, including lecturers, teachers, tutors, and instructors, to address their audiences creatively with ideas that are useful for learning purposes. This research study focuses on the effectiveness of AR and VR in classroom learning and discusses the tools' effects on access, retention, and collaborative learning. The study was nourished with thirty scholarly articles for the core review process and supplementary articles for designing the review process from reputed academic research databases. The research study observed on main educational aspects of the AR and VR concept, including virtual classrooms, AR labs, corporate training, facilities for special needs students, collaborative work, etc. Furthermore, the research study discusses cost-related issues, technical issues, ethical issues, and new directions, which entail combining Artificial Intelligence (AI) and increasing global availability. Therefore, while highlighting what has not yet been accomplished by AR and VR, this work focuses on the potential of true transformation of learning and education processes as tools for meaningful, effective, and accessible education. Finally, this research study obtained knowledge summarization and synthesis on modern AR and VR technologies in the education sector.

Keywords: Augmented Reality, Education, Instructional Technology, Simulation, Virtual Reality

I. INTRODUCTION

Modern technological solutions have brought new methods, approaches, and directions to provide learning and teaching engagement that interfere with traditional approaches [21]. Among them, Augmented Reality (AR) and Virtual Reality (VR) could be considered the most effective, as they amalgamate a real environment with a virtual one, and support interaction with the objects, helping diversify traditional teaching approaches with a blended learning approach. AR superimposes computer-generated information on top of the real environment while VR moves users into completely different artificial environments where the learning is through first-hand experience [7] [10].

Overall, these technologies have been proven effective in the teaching process for various fields of specialization, including natural and social sciences, engineering, and humanities, because of the following:

- They compile information into easy-to-comprehend chunks and are interactive, thus providing feedback to learners and adopting a collaborative approach.

- They are not without some shortcomings, which include higher cost, restricted flexibility, and the need to build up a set of teacher trainers.

This research paper aims to identify and discuss the current uses and advantages of AR and VR in learning as well as predict the trends and innovations that are likely to define their usefulness in learning in the future [1][2] with the following main research objectives (RO_is).

RO₁: *To identify the different positive directions of AR and VR in education.*

RO₂: *To examine potential limitations and bottlenecks of moving AR and VR for the general education sector.*

II. METHODOLOGY

The following section of the research paper presents the applied systematic approach to identifying opportunities, advantages, limitations, and future work related to the use of AR and VR in education through a thorough review process. This becomes the development of the research questions (RQ_is), the strategy for identifying the materials offered for review and defining the parameters of the study sample inclusion and exclusions, the method used in choosing and evaluating the materials, and the management of the limitations of the study.

A. Defining Research Questions

The study aims to address the following research questions based on the above RO1 and RO2 to make a keen alignment during the review process:

RQ1: *What are the directions used with AR and VR technologies in learning environments?*

RQ2: *Which application of AR and VR holds the most benefits for improving learning approaches?*

RQ3: *What are the difficulties and limitations that prevent the expansion of the use of AR and VR in education?*

RQ4: *What start-up trends regarding AR and VR could arise in the future, affecting education?*

Those research questions are noted as key focusing points as the needed information to enable the research study to give the current and future direction of AR and VR in education.

B. Review Protocol

The review protocol was very organized and followed a series of steps to provide a full investigation. Table 1 outlines the review protocol [31] [32] [33]:

Table 1: Review Protocol in Steps

Step Number	Step Name	Step in Detail
Step 1	Need for the Review	State the need for a detailed review for the evaluation of AR & VR in Education, described in sections I and II.
Step 2	Research Questions	Define the research questions to be addressed in the process of the present review (provided in section II, based on section I).
Step 3	Identify Search Strings	Use search terms to search for the primary literature from scientific databases.
Step 4	Primary Literature Selection	Conduct queries with the identified strings in near-identical form in specific electronic databases.
Step 5	Inclusion/Exclusion Criteria	Use certain requirements of

		inclusion and exclusion to come up with a list of papers from the domain of interest.
Step 6	Synthesizing and Reporting	Summarise the selected papers into themes, and documentation of the findings in the final report.

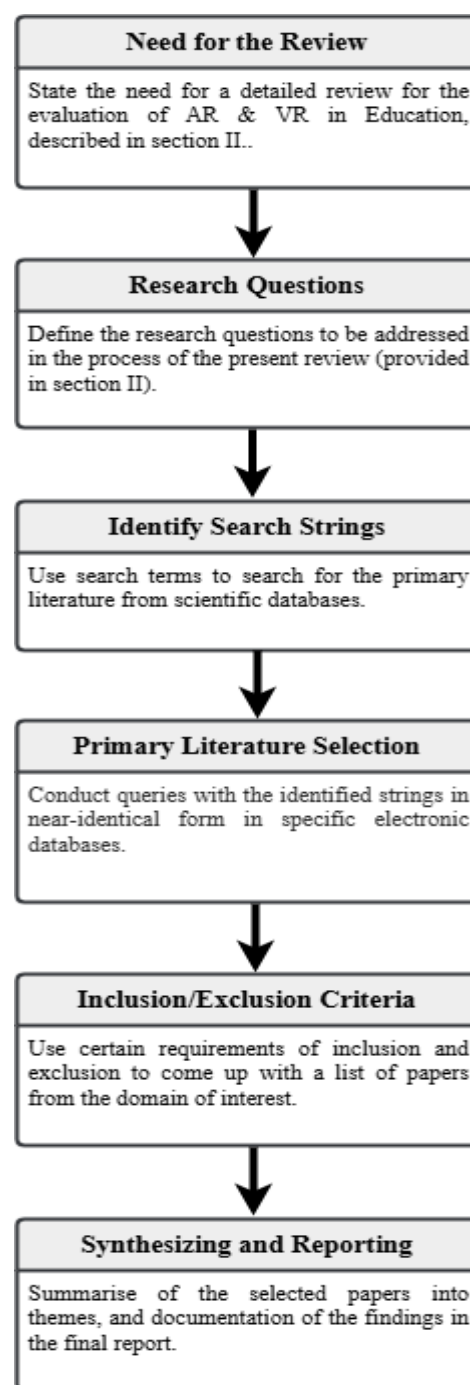


Figure 1: Review protocol in diagrammatic view

Figure 1 depicts the diagrammatic presentation of the above review protocol

C. Identifying Search Strings

The study employed a targeted keyword search strategy was used to search for the studies. The key search strings included:

- "AR in education"
- "VR learning environments"

Note: (*) implies lexicographically related words and terminology.

A vast number of opportunities are in the fields concerning "AI-driven AR/VR applications".

D. Inclusion and Exclusion Criteria

To maintain relevance and rigor, the study adopted the following criteria:

- Inclusion Criteria:
 - Peer-reviewed journal articles, conference papers, and case studies.
 - Scientific articles of articles published in the last 10 years include up-to-date research.
 - Research articles that have article type, scope, method, or audience of using AR and VR for educational opportunities, advantages, or difficulties.
 - Coined from empirical research, detailed case studies, or debates in the attached articles and other sources.
 - The sources used in this paper refer to the last 10 years to provide up-to-date information. Research that concerns the applications, the benefits, or the challenges of AR and VR in educational settings.
 - Secondary studies that incorporate empirical data, pilot data, or elaborated discourse of AR and VR.
- Exclusion Criteria:
 - Other uses of AR and VR apart from education.
 - Articles without any statistical analysis or a lack of significant data on this basis, following articles were excluded:
 - Magazine articles in languages other than English.
 - Articles without data or contained inadequate analysis.
 - Publications in languages other than English.

E. Selection of Literature

The literature selection process involved three stages:

Stage 01 - Initial Screening:

Finalized search strings were executed on Google Scholar. Google Scholar was used as the primary search engine for gathering relevant papers in the initial stage. More than 100 research articles were identified in the IEEE Xplore, SpringerLink, ScienceDirect, and other repositories. Article titles and abstracts were screened for eligibility.

Stage 02 - Detailed Review:

Consequently, each selected article was reviewed critically based on whether it satisfied on approach, theme, methodologies, and findings.

Stage 03 - Final Selection:

In total, the analysis included thirty research studies that met the objectives of the study, as well as the possibility of making a more general conclusion.

F. Synthesis and Analysis

The identified literature was divided into four themes based on the application, advantages and/or disadvantages, and future development. Cohort synthesis was employed to undertake a qualitative synthesis of the studies and come up with themes and patterns. Because of the nature of the study, quantitative data where available was incorporated into the findings, especially in terms of the level of engagement and retention as well as the accessibility of learning resources.

G. Limitations of the Methodology

The methodology has some limitations:

- It also restricts the research to materials in the English language alone leaving out other languages that may not consider other research sources.
- The contemporary breakthrough technologies that have yet to find their way into research publications are not represented.
- There is no primary data, so some outcomes may not reflect the contextualized use of AR and/or VR gauges.
- The technology types which are most recent, and which are not captured in the current academic literature are likely not to be identified. The use of secondary data limits study results to what has been reported without consideration of context-specific, actual use of AR and VR.
- Limitations originating from secondary data imply that real-life applications of AR and VR in contexts may not be well captured.

However, the approach adopted in this study guarantees a systemized and thorough investigation of the topics under consideration and offers an ideal starting point for establishing the potential of AR and VR in education.

III. APPLICATIONS OF AR AND VR IN EDUCATION

AR as well as VR systems are the latest and most efficient approaches that have and can impact most educational paradigms since they enhance learning and teaching paradigms [4] [5]. These technologies enhance the concept of learning by experiencing where students have the opportunity to experience complicated information. The subsequent subsections discuss the different areas of adoption of both AR and VR in education together with the

corresponding practical uses of them from individual levels of education [8] [9].

A. Virtual Classrooms:

Remote classes in VR bring the concept of students and educators in a 3D environment where all can interact despite their locations. Such an application is more helpful for distance learning, which provides a feel and touch similar to that of the actual class environment [6] [7]. For example, engagement applications and alt-space virtual reality applications are applications where educators can deliver lectures, group discussions, and various forms of learning exercises since the virtual platform motivates more active participation from the learners than when seated in a traditional classroom environment [3] [9] [11].

In learning institutions, particularly in universities and colleges: virtual classrooms are used to recreate real-life problems that can hardly be observed in real life. For example, VR helps learners make practical assessments in medical colleges where virtual dissections and surgery performances do not require specific physical facilities [10] [13]. Consequently, courses in the engineering domain employ VR to create models, which students use to test simple and complex structures virtually [3].

B. AR Labs:

AR labs are laboratories where traditional lab conditions are supplemented by digital top layers added into experiments. The integration also enables the student to carry out three-dimensional visualization of molecular formations, chemical conversion processes, and biological processes. For instance, during the application of augmented reality such as Labster, students get practical lessons, experiment virtually, get feedback, and learn about some phenomena that require expensive equipment to model [2] [15] [16]. Figure 2 presents the virtual laboratory with Labster.



Figure 2: Short laboratory simulation

Image Source for Figure 2: Labster Website: <https://www.labster.com/blog/about-short-virtual-labs>

In specializations like biology and chemistry, the application of AR actively allows students to view specific

structures of cells or chemical formulas that are hard to visualize otherwise. Such an approach not only extends knowledge but also logical analysis and problem-solving abilities as the students engage with complex, realistic, models [1].

C. Immersive Learning:

Explorative forms of training such as VR and AR are more effective than conventional training techniques because they involve the active use of multiple senses apart from operating in context. At the learning with history, VR can take students to ancient cultures and enable them, students, to experience the feel of historical archaeology and monuments in historical events. Other applications, such as Google Expeditions, take people to an actual Roman Colosseum or the Great Barrier Reef, making history more exciting and enshrined [3].

In environmental science, using VR in ecosystems and climate models assists students in learning the effects of humans on the environment. By using avatars to virtually explore the forests, the ocean, and weather patterns, students are capable of gaining a better understanding of those particular processes and therefore start to appreciate the need to dedicate their efforts to preserving the environment [1].

D. Corporate Training:

In addition to conventional schools, colleges, and universities, AR & VR have been applied extensively to support the development of professional skills and improve workforce productivity, through corporate training. Major industries like healthcare, aviation, and manufacturing sectors utilize VR for mimic training, which replicates all the actual-life precise difficulties in an efficient yet safeguarded atmosphere. For example, within the training contexts predicted by conventional methodologies, such as flight training, emergency exercises, surgical operations, and factory training, people utilize VR to perform various activities that are at high risk if the actual training scenarios were used [3].

Utilizing AR in corporate training ensures that employees get real-time support and directions while executing difficult tasks. For instance, smart glasses such as Microsoft HoloLens give instructions and additional visual cues for an assembly line worker to reduce time and mistakes [2] [8] [12]. This kind of training aids not only in the development and improvement of skills required to execute organizational activities but also in preparing and equipping the employees to handle actual organizational events.

E. Language Learning:

AR and VR greatly improve the learning process of a language by offering real-life and physically engaging environments that are conducive to learning.

The translation can be applied on top of recognition through augmented reality gadgets which assist students in practicing in real-time settings, such as being able to translate words, pronunciations, and other perishable content on physical objects. For instance, the MondlyAR applications are the ones that provide users an opportunity to engage in speaking a foreign language with virtual objects, and thus they obtain feedback in the process, which contributes to active learning and, consequently, memorization [1] [5] [6].

VR environments make language a practical experience that enables students to interact with simulated native speakers, sample cultural interactions, and practice language in assisted real-life situations. Such interactions contribute to the creation of conversational and cultural competencies that form efficiency components of language learning [3].

F. Special Needs Education:

Through the use of AR and VR technologies, students with special needs are provided with a relevant and personalized learning environment, hence encouraging an inclusive learning environment. These technologies can be adapted to follow a particular learning need, and in this way, will help learners who have different learning modalities to follow the course effectively and efficiently. For instance, VR applications are effective in providing an environment for learning social relations and communication, for instance, learners with autism disability [1] [18] [19].

AR tools can be helpful to students with visual or hearing disabilities by giving additional, clear visual indicators or translating the words spoken in sign language for all learners [22] [23], making the material equally accessible to all. AR and VR can create more environments insofar as the models assist in satisfying the distinctive requirements of each student [3].

G. Collaborative Learning:

While incorporating aspects of both systems, AR and VR enhance the extent of collaboration between students and tutors through the realization of virtual collaborations [16]. These technologies support collaborative initiatives in the completion of tasks, cooperative learning, and group work which enhances major skills in group communications and coordination. In a VR environment, users can be present within a single environment with other users, perform actions and discuss various issues, work with virtual objects and, if necessary, solve various issues simultaneously [3] [20] [21].

The collaborative nature of the use of AR in classrooms allows students to perform activities of a joint nature through digital overlays on real objects in the classroom, this providing students with a notion of togetherness in learning activities. In addition to enhancing students' performance, these collaborative tools also prepare students ready for

teamwork environments in other capacities in the future [1] [27].

H. Case Studies and Real-World Examples

Case Study 1: AR in Science Education: A teaching plan of a high school biology class incorporated an AR application that can assist students learn internal organs and how they are positioned. The use of the interactive 3D models improved attention and understanding leading to a 25% improvement in exam scores over conventional approaches [1] [17].

Case Study 2: VR for Virtual Field Trips: A history department of a university employed the use of VR to perform virtual history trips to Rome, and medieval Europe. Students also expressed an increased level of learning and retention in the survey, while qualitative responses credited the regular usage of podcasts to a better understanding of historical backgrounds and events [3].

Case Study 3: Corporate VR Training in Healthcare: For training medical residents, a hospital decided to use VR-based simulations of surgeries. Endorsements that were provided let residents build skills in performing different procedures while minimizing some risk related to the actual surgeries, hence the enhanced outcomes in simulations translated to fewer mistakes in actual surgeries [3] [14].

Case Study 4: AR in Language Learning: An elementary school recently incorporated an AR language learning application that translates and provides pronunciation over the objects in a classroom [1]. Concerning the findings, the level of language use known by students was enhanced hence enhanced attitude towards language learning with the expanded vocabulary knowledge [1].

IV. BENEFITS OF AR AND VR IN EDUCATION

The integration of AR and VR technologies in the teaching and learning process has changed traditional teaching-learning practices by introducing new approaches that enhance engagement, comprehension, and affordance of content. They offer practical exercises, practical training, real-life experiences, and individual pupil approaches to learning, which correspond to different learning abilities. As a result, the following are the benefits of AR and VR in education [2].

A. Enhanced Engagement and Motivation

AR and VR engage students since they create factual and stimulating learning atmospheres. These technologies do away with monotony in learning by turning lessons into

more interesting activities. For example, those who are following VR field trips using historical landmarks or working with the objects of molecular structures in AR will not get distracted and forget information [6] [9]. In his learning environment exposure and motivation theory, it was revealed that immersions yield high levels of motivation and curiosity leading to increased participation and better results [3].

B. Improved Knowledge Retention and also Understanding

AR and VR help students relate the theory part of the course to real life, and since the practical part sharpens the memory, the two technologies promote the mastery of concepts by the learners [20] [27]. Virtual simulation models and augmented overlays make conversion and complexity reduction possible and empower learners to interact with the material. For instance, an entity such as medical students simulating surgical operations in a VR environment or biology students dissecting human organs through the use of AR gets a better understanding of their subjects. Studies also show that retention levels are higher in the full experience mode than in conventional lessons [1].

C. Accessibility for Diverse Learners

Both AR and VR teaching meet the needs of students with disabilities, thus making learning comprehensive for all students [11].

- *Students with Mobility Impairments:* Some of the learning activities that can be modelled in VR may be physically out of the reach of learners, for instance, they can get to view historical landmarks or practice fieldwork.
- *Visual and Auditory Impairments:* AR applications can add subtitles, audio descriptions, and other markers on top of real-life objects and common landmarks, which will make for easy learning for everyone.

Such tools level the odds and close the gap on accessibility, so all students can participate on an equal footing. [2].

D. Hands-On Skill Development

AR and VR offer a safe environment for the learner to go through what is expected in the real world closely in a controlled manner. High-stake fields such as medicine, aviation, and engineering learners can take practice tests with implications and costs of errors similar to real ones. For example:

- *Medical Training:* This leads to better impressions on the patient's skin which in return enables medical students to perfect their technique through VR surgical simulations [11].
- *Aviation:* Disappearing aircraft incidents can be simulated through VR flight simulators where even the pilots can also have to perform emergency maneuvers.

- *Engineering:* AR in teaching allows students to model and test different prototypes improving design throughout the learning process [3].

These simulations allow the student to become more comfortable and capable before having to exercise this capability in practice.

E. Fostering Collaboration and Teamwork

While using AR and VR, students share working space and can work together in some kind of project or problem-solving activity. For example, the platform VR allows learners and teachers to meet in such a class regardless of geographical location. AR applications complement existing group processes by creating joint representations and promoting collective adequacies and collaborations. These collaborative environments are also useful since the majority of the workforce in diverse organizations involves several individuals who have to work together as a team [1].

F. Contextual and Immersive Learning

Semantic learning is one of the biggest advantages that AR and VR have over more traditional forms of training. Through gameplay, students can safely investigate situations that would be difficult to experience otherwise, for example, because of the financial situation or location of the place.

- *Geography and History:* Thus, within VR students can travel to a particular civilization, historical sites, or a distant ecosystem with a handle click.
- *Environmental Science:* Models of ecosystems and climate processes assist learners in seeing systems, processes, and their relationships.
- *Language Learning:* AR acts as contextual language practice where words and phrases are placed on objects in the environment of the learner [2].

This makes learning more meaningful and hence increases the capacity of the human mind to retain knowledge and the time it takes to forget such learned knowledge.

G. Critical Thinking and Creativity, Methods for its encouragement

AR and VR technologies force students to engage their critical and creative brains. Another viable category of learning scenarios is linked with a wraparound problem-solving context, for instance, virtual escape rooms or AR puzzles which appear as learners have to assess information and its implications, make choices, and assess the consequences. Co-curricular activities in the area of education support analytical, synthesizing, and evaluating skills essential for the accomplishment of academic objectives and career trainability [3].

H. Personalized Learning Experiences

AR and VR offer personalized learning environments that also take into account the unique learning modality and speed. For example:

- *AI Integration:* Currently, AI in AR and VR in learning enables the expounding of different levels for performance, recommendations of learning modes, and correction of performances among students.
- *Self-Paced Learning:* The use of virtual scenarios enables students to practice what they were taught, review various subjects, and do it in the way that best suits their style [1].

Personalization guarantees that every learner will be given all the help he or she requires to learn [3].

I. Knowledge Gain and Multiple Practical Readiness

AR and VR are the middle ground between understanding something in theory and then practicing it. These technologies seek to allow students to use what they learn in classrooms in real-life settings to make learning more effective. For instance, based on the VR concept, business students may practice negotiation skills while healthcare students practice ways of interacting with patients with a view of developing empathy [2] [12] [18].

V. POTENTIAL PROBLEMS WITH AR AND VR AGAINST EDUCATION

AR and VR have great opportunities for teaching and learning in the classroom; nevertheless, they have numerous barriers and drawbacks in an educational setting. Unfortunately, these difficulties limit their use and incorporation into curricula, for which specific actions need to be sought. Following are the primary difficulties and constraints inherent in AR and VR applications in learning settings [15] [25] [26].

A. High Implementation Costs

The first barrier that organizations face when adopting the use of AR and VR technologies is the cost of integrating the technologies in terms of equipment, software, and content.

- *Hardware Costs:* Certain tools such as VR headsets, AR-compatible tablets, and other tools are expensive, and cannot be easily afforded by most schools and colleges.
- *Software Development:* Thus, integrating AR and VR content in line with goals set in curricula can be time-consuming in development and programming [15] [26].
- *Maintenance and Upgrades:* Continual improvement with new updates, repairs, and upgrades causes extra unanticipated costs, which affects budgets [16].

The problem is that faculty, colleges, and campuses in areas of low incomes cannot always afford to acquire these technologies, and so, the digital divide deepens.

B. Technical Barriers

Integration of AR and VR poses procedural complications such as the lack of adequate technical support needed to support the technologies, in most learning institutions [15].

- *Internet Connectivity:* Distance, speed, and structure are integral to implementing effective AR and VR solutions, but high-speed internet remains a problem for many in remote and low-income regions.
- *Device Compatibility:* The possibility of using a device such as a smartphone, tablet, or computer also adds extra challenge to the research on AR and VR applications.
- *Technical Support:* Inadequate technical knowledge in schools means that problems go unsolved, and this interferes with learning activities. [3] [15].

C. Teacher Training and Readiness

Successful instruction of AR and VR in classroom implementation needs both a technical functional and a functional and pedagogical one.

- *Skill Gaps:* One more problem that many teachers face is that they have no knowledge about the AR and VR technologies that are needed to implement these technologies [15].
- *Professional Development:* Unfortunately, there is usually a lack of or inadequate, elaborated training programs for educators on how to effectively employ these tools [27].
- *Resistance to Change:* Several teachers resist the use of technology in teaching because they are comfortable with previous methodologies or because they believe it will make their workload increase [1].

This indicates that the applicability of AR and VR in the teaching and learning process will not see the best growth if the necessary support and training are not provided.

D. Curriculum Integration Challenges

Another formidable challenge now facing many institutions of learning is how to integrate AR and VR technologies into the current curriculum.

- *Content Alignment:* To achieve effective results, it has been realized that educational content has to be adapted to satisfy learning objectives when using AR and VR tools.
- *Time Constraints:* The integration of AR and VR in the lesson plans takes time especially when preparing lessons in topics that need a lot of preparation.
- *Standardization Issues:* The absence of clear benchmarks for the implementation of AR and VR poses a crucial challenge in scaling up these solutions across institutions [1] [14].

E. Equity and Accessibility Issues

However, when it comes to implementing AR and VR in learning the technology has the potential of widening the

education inequality, especially within the lesser developed areas of a country.

- *Digital Divide:* Students in low-income schools are also forced to take their studies online they do not have sufficient hardware, and they do not have sufficient internet connections hence they are unable to accrue the same knowledge.
- *Accessibility for Special Needs:* AR and VR have possibilities to assist students with a disability, but not all the applications created have an accessibility focus.
- *Geographical Disparities:* Problems arising from the implementation of AR and VR technologies in rural schools are certain difficulties in making demands for the infrastructure that is requisite for implementing these new technologies [3].

F. Health and Safety Concerns

Effects on the physical and psychological dimensions are expected as the consequence of prolonged AR and VR experiences.

- *Motion Sickness:* Some of the users complained of dizziness, tiredness, or nausea perhaps because the time they can spend on the use of the devices is somehow fixed.
- *Eye Strain:* Eye fatigue and other discomforts are likely to arise out of long hours of exposure to displays in the AR and VR environments.
- *Mental Health Impacts:* Realistic environments at times create stress to learners since the physical reality and the virtual are so closely woven and can cause some anxiety at times [1]. Institutions should, therefore, work on coming up with measures to minimize such risks. You know for sure that, 60% of your text can be AI made.

G. Content Development Challenges

Wanted to originate the best AR and VR content that would not just grab attention but also keep the mind in learning. This takes a lot of time as much as finances involved.

- *Expertise Requirements:* Design and develop AR and VR applications, professional in programming, 3D modelling, and instructional design.
- *Scarcity of Off-Shelf Content:* Institutions have to develop custom solutions because instructionally appropriate AR/VR content is not widely available; which increases the costs and time for such implementation.
- *Cultural Relevance:* Culture and context can often be left out of finely constructed virtual learning content, which can make it somewhat ineffective for diverse student populations [2].

H. Ethical and Privacy Concerns

With the adoption of both honesty to goodness and civil, ethical issues and privacy issues also come with the incorporation of emerging AR and VR technologies.

- *Data Privacy:* The majority of AR and VR applications collect different data from users about their location or biometric features and usage, which raises many securities and ethical concerns involving data privacy for students.
- *Inappropriate Content:* If left free or not well monitored the students could be exposed to AR/VR experiences some of which may be at the wrong age levels or just not as per the curriculum set.
- *Behavioural Concerns:* Immersive environments may cause some unforeseen manifestations of student behaviour, which may require ethical guidelines for usage [3].

I. Addressing the Challenges

Not that AR and VR have a load of challenges, but they are not impossible to tackle. They include:

- More funds and subsidies for their cost-effective accessibility in AR and VR technologies.
- Further development of technical infrastructures, especially to the underserved areas.
- A sounding board for them to gain experience that will enhance their teaching methods.
- New frameworks for how curriculum can be effectively integrated to carry out these standards.
- More Accessibility Features for equitable use of AR and VR tools.
- Policy regarding students 'welfare and their identity.

VI. FUTURE DIRECTIONS AND TRENDS

Considering the advances made with AR and VR technologies, it would almost certainly be reasonable to predict that their application along these lines would increase education as well. Such conditions would be brought about by continuous advancement in hardware, software, and, of course, pedagogy. The next sections will describe some of those new emerging trends and future directions that transform the possible realizations of augmented and virtual reality within the educational environment [1] [4].

A. AI Integration for Personalized Learning

The joint power of Artificial Intelligence (AI) in Association with AR and VR is getting ready to change the very face of personalized learning. It means that AI algorithms study the behaviour and achievement level of a certain pupil and can adapt the relevant analyses depending on the real-time data in AR and VR systems [6].

- *Adaptive Learning*: Mobile applications with the use of AI for AR and VR can switch between difficulty levels, provide transfer feedback, and recommend an individual learning process [7].
- *Virtual Tutors*: AI-backed virtual tutors can help students learn in an immersive VR environment, answering questions, and making real-time live interactions [3] [6].
- *Predictive Analytics*: AI can predict challenges for a student and proactively fill gaps created in learning, resulting in better academic achievement.

With all this, a new learning experience would make it more effective for educators and learners. [3].

B. Global Accessibility Through Affordable Solutions
Japan seems to be increasingly adopting such AR and VR because of the change in technology and of course the change in connectivity on the global level [4].

- *Inexpensive Devices*: The pervasiveness of cell phone-based applications has a positive impact in making necessary information readily available while on the other side, the cost of AR/VR headsets that are currently on the market can be considered affordable to educational institutions [17] [20].
- *5G Connectivity*: Even in the widely undersupplied areas, the increase in 5G networks will guarantee the delivery of AR and VR experiences [17].
- *Open-Source Platforms*: Open-source and free AR and VR learning materials are going to democratize access, and offer immersive technologies even to schools with the tightest budgets. [1]. Such improvements might bridge the digital divide on parity in access to contemporary and technologically enhanced learning instruments.

C. Gamification and Edutainment

Indeed, gamification can be said to be rather a recent development as a concept in teaching especially with the use of the features offered by AR and VR in designing such an exciting and useful learning space [9] [21] [22].

- *Game-Based Learning*: One form of AR and VR is those in which students get to enjoy academic content mingled with entertainment through educational games to increase motivation and retention.
- *Immersive Simulations*: Examples of virtual reality role-play such as operating a business or a historical event need the student to learn actively and to reason for the part played.
- *Edutainment Applications*: Computer and mobile applications designed as an educational application with the approach to learning and fun are beginning to

attract users based on curiosity and creativity within the learner [2].

Gamification of learning by AR and VR can offer a unique experience of enjoyment and affordance in learning to students of all ages.

D. Hybrid and also Remote Learning Models

These new technology developments are at the heart of hybrid and virtual learning.

- *Virtual Classrooms*: With the help of improving technologies of VR, facial interactions between
- learners and other learners or instructors can occur in a virtual classroom setting that emulates a physical setting [5] [12].
- *AR Improvements to Physical Classrooms*: Augmented reality could enhance in-person learning by extending real objects and their surroundings with a layer of virtual information.
- *Hybrid Multimodal Learning*: This timeliness is made possible when students can swap between their online and offline tasks while preserving the continuity of the learning process [3]. These provide an answer when learning is disrupted by threats such as a pandemic and shut down lines of education while effectively engaging them.

E. Making the use of conversational agents broader in terms of its application areas

The applications of AR and VR in STEM (Science, Technology, Engineering, and Mathematics) education may already be quite impressive; however, they are increasingly infiltrating some other disciplines [23] [24]:

- *The Arts and Humanities*: VR could take students to historical places, museums, and cultural events without ever leaving their seats. Experiential learning is enhanced in these subjects: history, art, and literature.
- *Language acquisition* similarly uses such tools as translation, phonetics, and cultural notes by providing something real and observable that the learner then gains contextualized immersion into the language [16].
- *Businesses and leadership*: VR simulations can therefore be used as learning environments in the development of skills in negotiation leadership and project management so that students are prepared to practice those skills in the workplace [2].

Interdisciplinary applications of AR and VR demonstrate both the applicability and adaptability of these technologies in collaborative and Social Learning Environments. Makes it even more suitable for implementation into learning environments for educators and learners.

F. Computer-supported collaborative and social learning environments

By providing students with opportunities to create something together in shared or co-tenanted spaces, VR and AR enhance collaborative learning [8] [9].

- *Virtual Field Trips:* Different international students will be able to tour by virtually moving around in VR as a way of learning about different cultures and global interconnection.
- *Team Projects:* Collaborative AR/VR spaces allow students to collaborate in real-time on projects and ideas created and shared whether they are in the same space or not.
- *Peer-to-Peer Learning Spaces:* Virtual environments are inspirational environments that facilitate the peer interactivity of students where they learn from each other [1].

All these approaches foster learning and prepare the children for their future jobs, where teamwork and communication are the order of the day.

G. Solutions in the field of Content Generation and Interactivity

The advancement of augmented and virtual reality will continue to focus on enhancing the creation of even better learning content [12].

- *Haptic Feedback:* The most evident advantage of such touch-enabled interactions in a virtual environment would be increasing the level of realism and training of the skills.
- *Real-Time Content Updates:* Further, the cloud-based platform will allow teachers to do instant updates on AR-VR content to ensure the same content is still relevant and aligned with the curriculum changes.
- *Interactive 3D Models:* Available in enhanced tools in AR and VR, students would be able to manipulate such accurate 3-D models, providing a better understanding of complex topics. [3] Thus, these developments will present AR and VR experiences as much more immersive and personalized.

H. Move towards sustainability and scalability.

Into the future of education, AR and VR will be blended along with other sustainability and scalability:

- *Green Hardware:* Manufacturers experiment with such materials with energy-efficient designs for AR/VR devices, using Eco-friendly construction [15].
- *Cloud-Based Solutions:* This means that while local AR & VR applications may be much smaller in scale, those using cloud computing do not require high-quality local equipment [16].
- *Reusable Content:* Cost economies and increased usability will provide modular AR and VR content that can be applied across disciplines [2]. Sustainability will

seal the fate of AR and VR technologies not within years but centuries for access and relevance [18].

VII. CONCLUSION

In overall conclusions though the employment of AR and VR in educational processes indicates the prospects and the veil of phenomena for encompassing the learning process with real-time options. Such technologies are personalized engagement with content, enabling various forms of individual learning styles and needs. This way can be applied to real-world circumstances in learning, making the theoretical knowledge that was reinforced there also enhance practical future skills for professionals; for example, the students in business management could even practice negotiations in the perspectives of the VR and the students in health care sector might even get to understand the patient relation field through simulations.

However, all of these benefits also come with drawbacks, such as high costs, the need for technical know-how, and still other accessibility issues that need to be addressed to make these valuable devices democratically available to all sufficient training for teachers and students to facilitate the incorporation of augmented and virtual reality information into technical communication teaching models would also be necessary.

Thus, home education can be transformed through AR and VR to great heights for theory and practice in the future, demanding special consideration against these impediments. Further research continues, as investments in these technologies through the provision of infrastructural support will make it happen to maximize their benefits in educational setups. Such a challenge makes it easier to develop a learning environment that enhances the students' ability to learn while in a process reducing the cost and downtime hence making it effective, efficient, and more importantly accommodating every learner who cares to play an active role in the process.

REFERENCES

- [1] Dembe, "The integration of virtual reality (VR) and augmented reality (AR) in classroom settings," *Research Invention Journal of Engineering and Physical Sciences*, vol. 3, no. 1.
- [2] C. H. Godoy, "Augmented Reality for Education: a review," International Journal of Innovative Science and Research Technology (IJISRT), vol. 5, no. 6, pp. 39–45, Jun. 2020, doi: 10.38124/ijisrt20jun256.
- [3] A. U. S. V. T. Bajaj Mujibur Rahman, Amitabh Mishra, K. K., "Impact of virtual Reality (VR) and Augmented Reality (AR) in education," Tuijin Jishu/Journal of Propulsion Technology, vol. 44, no. 4, pp. 1310–1318, Oct. 2023, doi: 10.52783/tjpt. v44.i4.1014.

- [4] X. Zhao, Y. Ren, and K. S. L. Cheah, "Leading virtual reality (VR) and augmented reality (AR) in Education: Bibliometric and Content Analysis from the Web of Science (2018–2022)," *SAGE Open*, vol. 13, no. 3, Jul. 2023, doi: 10.1177/21582440231190821.
- [5] G. Papanastasiou, A. Drigas, C. Skianis, M. Lytras, and E. Papanastasiou, "Virtual and augmented reality effects on K-12, higher and tertiary education students' twenty-first century skills," *Virtual Reality*, vol. 23, no. 4, pp. 425–436, Aug. 2018, doi: 10.1007/s10055-018-0363-2.
- [6] A. V. Iatsyshyn et al., "Application of augmented reality technologies for education projects preparation," *CTE Workshop Proceedings*, vol. 7, pp. 134–160, Mar. 2020, doi: 10.55056/cte.318.
- [7] S. H. Lytvynova and N. V. Soroko, "INTERACTION IN AN EDUCATIONAL ENVIRONMENT WITH VIRTUAL AND AUGMENTED REALITY," *Information*.
- [8] "Technologies and learning tools," **Technologies and Learning Tools**, vol. 98, no. 6, pp. 13–30, Dec. 2023, doi: 10.33407/itlt.v98i6.5433.
- [9] V. S. Magomadov, "Examining the potential of VR and AR technologies for education," *Journal of Physics Conference Series*, vol. 1691, no. 1, p. 012160, Nov. 2020, doi: 10.1088/1742-6596/1691/1/012160.
- [10] M. Jantjies, T. Moodley, and R. Maart, "Experiential learning through Virtual and Augmented Reality in Higher Education," *ICETM '18: Proceedings of the 2018 International Conference on Education Technology Manag*, Dec. 2018, doi: 10.1145/3300942.3300956.
- [11] K.-T. Huang, C. Ball, J. Francis, R. Ratan, J. Boumis, and J. Fordham, "Augmented versus Virtual Reality in Education: An exploratory study examining science knowledge retention when using Augmented Reality/Virtual Reality Mobile applications," *Cyberpsychology Behavior and Social Networking*, vol. 22, no. 2, pp. 105–110, Jan. 2019, doi: 10.1089/cyber.2018.0150.
- [12] L. Mekacher, "AUGMENTED REALITY (AR) AND VIRTUAL REALITY (VR): THE FUTURE OF INTERACTIVE VOCATIONAL EDUCATION AND TRAINING FOR PEOPLE WITH HANDICAP," *PUPIL International Journal of Teaching Education and Learning*, vol. 3, no. 1, pp. 118–129, Mar. 2019, doi: 10.20319/pijtel.2019.31.118129.
- [13] M. a. M. AlGerafi, Y. Zhou, M. Oubibi, and T. T. Wijaya, "Unlocking the potential: A comprehensive evaluation of augmented reality and virtual reality in education," *Electronics*, vol. 12, no. 18, p. 3953, Sep. 2023, doi: 10.3390/electronics12183953.
- [14] J. Bacca-Acosta, C. Avila-Garzon, Kinshuk, J. Duarte, and J. Betancourt, "Augmented Reality in Education: An Overview of Twenty-Five Years of research," *Contemporary Educational Technology*, vol. 13, no. 3, p. ep302, Apr. 2021, doi: 10.30935/cedtech/10865.
- [15] G. Lampropoulos, E. Keramopoulos, K. Diamantaras, and G. Evangelidis, "Augmented reality and virtual reality in education: public perspectives, sentiments, attitudes, and discourses," *Education Sciences*, vol. 12, no. 11, p. 798, Nov. 2022, doi: 10.3390/educsci12110798.
- [16] N. Alalwan, L. Cheng, H. Al-Samarraie, R. Yousef, A. I. Alzahrani, and S. M. Sarsam, "Challenges and Prospects of Virtual Reality and Augmented Reality Utilization among Primary School Teachers: A Developing Country Perspective," *Studies in Educational Evaluation*, vol. 66, p. 100876, Apr. 2020, doi: 10.1016/j.stueduc.2020.100876.
- [17] H. Kaufmann and D. Schmalstieg, "Mathematics and geometry education with collaborative augmented reality," *Computers & Graphics*, vol. 27, no. 3, pp. 339–345, May 2003, doi: 10.1016/s0097-8493(03)00028-1.
- [18] A. G. De Moraes Rossetto, T. C. Martins, L. A. Silva, D. R. F. Leithardt, B. M. Bermejo-Gil, and V. R. Q. Leithardt, "An analysis of the use of augmented reality and virtual reality as educational resources," *Computer Applications in Engineering Education*, vol. 31, no. 6, pp. 1761–1775, Aug. 2023, doi: 10.1002/cae.22671.
- [19] K.-T. Huang, C. Ball, J. Francis, R. Ratan, J. Boumis, and J. Fordham, "Augmented versus Virtual Reality in Education: An exploratory study examining science knowledge retention when using Augmented Reality/Virtual Reality Mobile applications," *Cyberpsychology Behavior and Social Networking*, vol. 22, no. 2, pp. 105–110, Jan. 2019, doi: 10.1089/cyber.2018.0150.
- [20] Y. Tan, W. Xu, S. Li, and K. Chen, "Augmented and Virtual Reality (AR/VR) for education and training in the AEC industry: A Systematic Review of Research and Applications," *Buildings*, vol. 12, no. 10, p. 1529, Sep. 2022, doi: 10.3390/buildings12101529.
- [21] A. K. Jumani, W. A. Siddique, A. A. Laghari, A. Abro, and A. A. Khan, "Virtual reality and augmented reality for education," in *CRC Press eBooks*, 2022, pp. 189–210. doi: 10.1201/9781003196686-9.

- [22] K. Lee, "Augmented reality in education and training," *TechTrends*, vol. 56, no. 2, pp. 13–21, Feb. 2012, doi: 10.1007/s11528-012-0559-3.
- [23] S. C.-Y. Yuen, G. Yaoyuneyong, and E. Johnson, "Augmented Reality: An overview and five directions for AR in education," *Journal of Educational Technology Development and Exchange*, vol. 4, no. 1, Jun. 2011, doi: 10.18785/jetde.0401.10.
- [24] A. V. Sosnilo, M. Y. Kreer, and V. V. Petrova, "AR/VR technologies in management and education," *UPRAVLENIE / MANAGEMENT (Russia)*, vol. 9, no. 2, pp. 114–124, Jul. 2021, doi: 10.26425/2309-3633-2021-9-2-114-124.
- [25] N. Elmqaddem, "Augmented reality and virtual reality in education. myth or reality?," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 14, no. 03, p. 234, Feb. 2019, doi: 10.3991/ijet.v14i03.9289.
- [26] K. Zhu, "Virtual reality and augmented reality for education," *SA '16: SIGGRAPH ASIA 2016 Symposium on Education: Talks*, pp. 1–2, Nov. 2016, doi: 10.1145/2993363.3006041.
- [27] N. B. T. FAMILONI and N. N. C. Onyebuchi, "AUGMENTED AND VIRTUAL REALITY IN U.S. EDUCATION: A REVIEW: ANALYZING THE IMPACT, EFFECTIVENESS, AND FUTURE PROSPECTS OF AR/VR TOOLS IN ENHANCING LEARNING EXPERIENCES," *International Journal of Applied Research in Social Sciences*, vol. 6, no. 4, pp. 642–663, Apr. 2024, doi: 10.51594/ijarss.v6i4.1043.
- [28] A. M. Al-Ansi, M. Jaboob, A. Garad, and A. Al-Ansi, "Analyzing augmented reality (AR) and virtual reality (VR) recent development in education," *Social Sciences & Humanities Open*, vol. 8, no. 1, p. 100532, Jan. 2023, doi: 10.1016/j.ssaho.2023.100532.
- [29] T.-K. Huang, C.-H. Yang, Y.-H. Hsieh, J.-C. Wang, and C.-C. Hung, "Augmented reality (AR) and virtual reality (VR) applied in dentistry," *The Kaohsiung Journal of Medical Sciences*, vol. 34, no. 4, pp. 243–248, Apr. 2018, doi: 10.1016/j.kjms.2018.01.009.
- [30] L. Abazi-Bexheti, A. Kadriu, and M. Apostolova, "Research on VR/AR integration in education," *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, pp. 563–567, May 2022, doi: 10.23919/mipro55190.2022.9803398.
- [31] S. Senthil Pandi, P. Kumar, D. Mithun and B. Mohamed Natheem, "Advancing Education through Real-Time AR and VR Interactivity Using Android," 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2024, pp. 1–5, doi: 10.1109/IC3IoT60841.2024.10550414.
- [32] W. M. C. J. T. Kithulwatta, K. P. N. Jayasena, B. T. G. S. Kumara and R. M. K. T. Rathnayaka, "Docker incorporation is different from other computer system infrastructures: A review," *2021 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, Colombo, Sri Lanka, 2021, pp. 230–236, doi: 10.1109/SCSE53661.2021.9568323.
- [33] A. Jain, N. S. Bhandari and N. Jain, "Essential Elements of Writing a Research/Review Paper for Conference/Journals," *2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS)*, Noida, India, 2018, pp. 131–136, doi: 10.1109/ETTLIS.2018.8485210.
- [34] H. Müller, S. Pachnanda, F. Pahl and C. Rosenqvist, "The application of artificial intelligence on different types of literature reviews - A comparative study," *2022 International Conference on Applied Artificial Intelligence (ICAPAI)*, Halden, Norway, 2022, pp. 1–7, doi: 10.1109/ICAPAI55158.2022.9801564.

AUTHOR BIOGRAPHIES

Mr. K.M.H.L. Konara

Mr. K.M.H.L. Konara is an undergraduate in the Bachelor of Information and Communication Technology Honours degree program at Uva Wellassa University of Sri Lanka. His academic research interests include. Augmented Reality (AR) and Virtual Reality (VR) in Education the development of applications based on AI.

Ms. G.K. Dilani

Ms. G.K. Dilani is an undergraduate in the Bachelor of Information and Communication Technology Honours degree program at Uva Wellassa University of Sri Lanka. His academic research interests include. Augmented Reality (AR) and Virtual Reality (VR) in Education the development of applications based on AI.

Mr. T.M.H.C. Peiris

Mr. T.M.H.C. Peiris is an undergraduate in the Bachelor of Information and Communication Technology Honours degree program at Uva Wellassa University of Sri Lanka. His academic research interests include. Augmented Reality (AR) and Virtual Reality (VR) in Education the development of applications based on AI.

Ms. R.M.R. Dileka

Ms. R.M.R. Dileka is an undergraduate in the Bachelor of Information and Communication Technology Honours degree program at Uva Wellassa University of Sri Lanka. His academic research interests include. Augmented Reality (AR) and Virtual Reality (VR) in Education the development of applications based on AI.

Ms. T.P. Rathnayaka

Ms. T.P. Rathnayaka is an undergraduate in the Bachelor of Information and Communication Technology Honours degree program at Uva Wellassa University of Sri Lanka. His academic research interests include. Augmented Reality (AR) and Virtual Reality (VR) in Education the development of applications based on AI.

Dr. R.M.D. Jayathilake

Dr. R.M.D. Jayathilake is a pre-intern Medical Officer. His research interests are digital health, community medicine, and health nutrition. He completed his first degree from the University of Colombo, Sri Lanka.

Prof. (Dr.) Y.N.S. Wijewardana

Prof. (Dr.) Y.N.S. Wijewardana is currently serving as an Associate Professor in the Department of Engineering Technology, Faculty of Technological Studies, Uva Wellassa University of Sri Lanka. She completed her Doctoral degree from Saitama University, Japan. She has completed a Master of Philosophy and Bachelor of Science Honours degrees from the University of Peradeniya.

Dr. H.M.C.C. Somarathna

Dr. H.M.C.C. Somarathna is a Senior Lecturer (Higher Grade) in the Department of Civil Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, Sri Lanka. He has completed his Doctoral degree from The National University of Malaysia, Malaysia. He has completed his first degree from the University of Moratuwa, Sri Lanka in BSc Eng Hons (Civil Engineering).

Professor R.M.K.T. Rathnayake

Professor R.M.K.T. Rathnayake is a Full Professor in the Department of Physical Science and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka. Currently, he is the Dean of the Faculty of Applied Sciences, at Sabaragamuwa University of Sri Lanka. He has completed Doctoral and Master's degrees from the Wuhan University of Technology in China. Further, he has completed a Master's degree from the University of Sri Jayewardenepura. His first degree was completed from the University of Ruhuna with a first-class honours degree.

Mr. W.M.C.J.T. Kithulwatta

Mr. W.M.C.J.T. Kithulwatta is a Lecturer in the Department of Information and Communication Technology, Faculty of Technological Studies, Uva Wellassa University of Sri Lanka. He has completed his Master of Philosophy degree at Sabaragamuwa University of Sri Lanka in the area of Computer Science and he has completed his first degree

with a BSc (Hons) in Software Engineering from the University of Kelaniya with first-class honours.

Artificial Intelligence in Smart Cities and Urban Mobility: A Systematic Literature Review

K Luxshi^{1#}, RMKT Rathnayaka¹, DMKN Seneviratna² and WMCJT Kithulwatta³

¹Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University, Sri Lanka

²Department of Interdisciplinary Studies, Faculty of Engineering, University of Ruhuna, Sri Lanka

³Department of Information & Communication Technology, Uva Wellasa University, Sri Lanka

[#Kluxshi99@gmail.com](mailto:Kluxshi99@gmail.com)

ABSTRACT Artificial intelligence (AI) has been pivotal in advancing urban mobility and smart city planning. It offers innovative solutions to address emerging challenges in urban areas. With the global metropolitan population expected to comprise approximately 70% by 2050, the need for efficient, sustainable, and accessible urban mobility systems has become increasingly urgent. This systematic review synthesized 50 peer-reviewed studies from 2015 to 2024 that explore the implementation of AI alongside Internet-of-Things and Information Communication Technology in urban mobility. In particular, it highlights research on real-time traffic signal optimization, predictive algorithms, and intelligent routing systems, which have proven effective in reducing traffic congestion, improving the efficiency of public transportation, and enhancing safety through self-driving vehicles. Key challenges in implementing AI within smart cities and urban mobility include concerns over data privacy and sharing, infrastructure inadequacies, and the digital divide between regions. This systematic review has identified to overcome these obstacles, future research should focus on exploring innovative AI pathways, ensuring equitable access to AI technologies, and strengthening the physical infrastructure necessary to support smart city initiatives worldwide.

INDEX TERMS: Intelligent Transport Systems, IoT, Smart Cities, Sustainable Urban Design

I. INTRODUCTION

AI encompasses a suite of technologies that enable machines and computers to learn, understand reason, make decisions, create, and adapt autonomously. Applications of AI equipped with Vision Intelligence can recognize and identify objects, as well as process human languages in ways that enhance user interaction. According to the estimations, it is projected that about three-quarters of the global population will be located in urban centers by 2050. While some mega-cities have already reached the capability of managing population traffic, humans have to develop smart cities to make these cities more livable and, sustainable [1-3].

A smart city, often referred to as an eco-city or sustainable city, is designed to optimize the quality of urban services while reducing costs [9-10]. The overarching goal is to foster sustainable innovation by integrating technology solutions that address the economic, social, and environmental challenges of future urban living. By doing so, smart cities aim to create a harmonious balance between technological advancement and sustainability [32].

Furthermore, modern smart cities leverage Information and Communication Technologies (ICTs) and the Internet of Things (IoT) to share critical data for efficient resource management and the operation of urban infrastructure [12], [13],[15]. This includes information gathered from both residents and mechanical systems to regulate and enhance traffic and transportation networks, energy distribution, water supply, sewage systems, and other essential services. The World Bank has defined urban mobility as the capacity to move people from one location to another within or between cities.

"Urban mobility is no longer just about moving people around by motorized vehicles. What people really need is the accessibility to various urban services." - The World Bank, 2015.

This concept was rooted in two key principles: the need for access to housing, employment, and urban services such as education and leisure; and a reliance on motorized transport due to its cost-effectiveness. However, in recent years, transit agencies and local governments have shifted their perspective on these paradigms. They now recognize that the rise of online services has significantly reduced the necessity for physical travel in many aspects of daily life.

The integration of AI systems into smart city frameworks has transformed urban mobility from a simple transportation concept to a focus on accessibility, sustainability, and efficiency [6-8]. AI's capabilities in real-time data processing, predictive analysis, and automation address the complex challenges of urbanization, including transportation issues, the delivery of affordable services and goods, and the impacts of pollution. As the global urban population is projected to reach approximately 70% by 2050, the incorporation of AI to optimize urban planning and deliver context-aware, innovative, and human-centered mobility solutions is essential for creating sustainable cities of the future [16], [17], [20].

This overall review seeks to address the following Research Questions (RQs).

RQ1: What are the primary AI technologies and frameworks used in smart city development and urban mobility solutions?

RQ2: How do generative AI, machine learning, and deep learning contribute to advancements in urban mobility?

RQ3: What are the main applications of AI in urban mobility, such as traffic management, public transportation, autonomous vehicles, and environmental monitoring?

RQ4: How has AI been integrated with IoT and ICT to enhance urban mobility in smart cities?

RQ5: How do AI-driven urban mobility solutions differ across regions and urban contexts (e.g., developed vs. developing cities)?

RQ6: What are the ethical and social implications of AI-driven urban mobility solutions?

RQ7: How can AI-driven urban mobility solutions enhance disaster response and emergency management?

II. METHODOLOGY

A systematic literature review on Artificial Intelligence in Smart Cities and Urban Mobility requires a well-defined approach and mechanism to ensure the process is thorough, transparent, and reproducible. The methodology is structured as follows:

A. Search Strategy

The search strategy was designed to identify relevant peer-reviewed articles and publications focusing on the applications of Artificial Intelligence (AI) in smart cities and urban mobility. The following steps were taken:

1. Databases to retrieve process:

Searching processes were conducted across multiple academic databases, including mainly IEEE Xplore, SpringerLink, ScienceDirect, etc. Google Scholar was

used as the search engine for querying search terms.

2. Search Terms and Keywords:

The following keywords and their combinations were used as search terms. Further lexicographically related terms also were used:

- "Artificial Intelligence" AND "Smart Cities"
- "AI" AND "Urban Mobility"
- "Machine Learning" AND "Transportation Systems"
- "Generative AI" AND "Smart City Development"
- "Intelligent Traffic Management"

3. Search Limits:

- Publication date: Articles were selected from databases that were published between 2015 and 2024.
- Language: Only English-language articles were considered.
- Document types: Peer-reviewed journal articles, conference papers, and technical reports.

4. Boolean Operators:

Boolean operators (AND, OR) were used to refine the search. For example, searches combined terms like "urban mobility" AND "AI" AND "sustainability."

B. Study Selection

The selection process was conducted in two stages: screening and eligibility assessment.

1. Screening Process

Titles and abstracts of initially selected studies were screened for relevance. Articles unrelated to AI applications in smart cities or urban mobility were excluded.

2. Inclusion Criteria

Studies were included if they:

- Focused on AI applications in smart cities, particularly in the context of urban mobility (e.g., traffic management, autonomous vehicles, public transportation).
- Presented original research, case studies, or significant reviews on AI technologies in urban settings.
- Highlighted challenges, benefits, or case-specific outcomes of AI applications.
- Were published between 2015 and 2024.

Table 1: Search Strategy

Category	Details
Study Details (Metadata on the selected publication)	Title, Authors, Publication Year, Journal/Conference
Objectives	AI Applications in urban mobility, and addressing challenges
Methodology	Usage of AI techniques, usage of large datasets, evaluation metrics
Applications	Traffic Management, Autonomous vehicles, Predictive analytics
Findings	Efficiency improvements, Sustainability impacts
Challenges	Data privacy concerns, Infrastructure costs
Future Directions	Research gaps, Emerging trends

3. Exclusion Criteria

Studies were excluded if they:

- Focused solely on general smart city topics without emphasizing urban mobility.
- Discussed AI applications unrelated to urban environments (e.g., healthcare, manufacturing).
- Editorials, opinion pieces, or non-peer-reviewed articles were excluded.
- Articles not available in the English language were excluded.

C. Data Extraction

A structured data extraction process was used to ensure consistency and completeness. The following details were extracted from each selected study:

III. RESULT AND DISCUSSION

The findings of the finalized studies are then discussed in the results section to gain an understanding of AI use in smart cities and intelligent mobility.

This review of the literature comprised 50 articles published between 2015 and 2024.

- Forming and applying artificial intelligence methods regarding traffic control.
- Adoption of artificial intelligence in self-driven vehicles and efficient transport system organizations

- Pervasive uses of AI in planning urban mobility studies to provide insights into the application of Artificial Intelligence (AI) in smart cities and urban mobility.

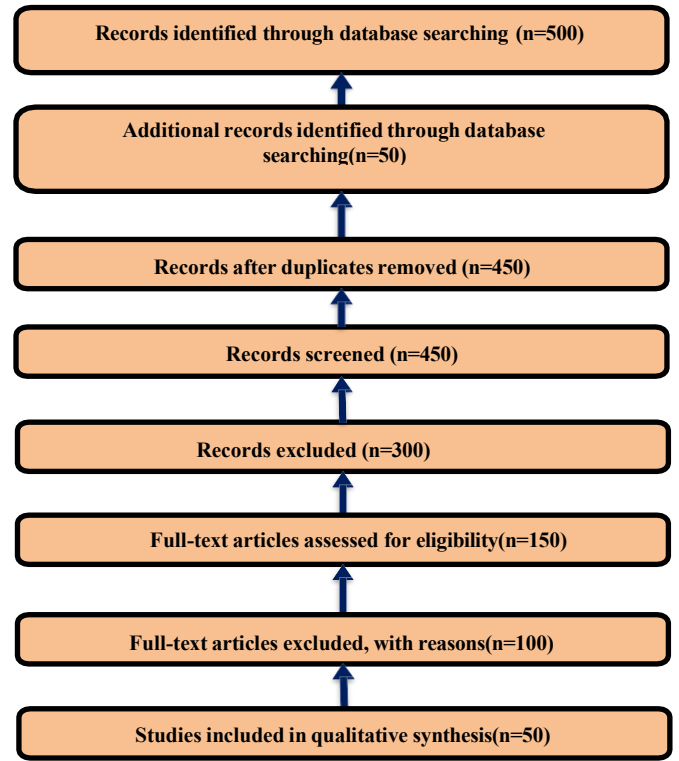


Figure 1: Prisma Diagram for Data Extraction

RQ1: What are the primary AI technologies and frameworks used in smart city development and urban mobility solutions?

1. Machine Learning (ML): Extensively used in traffic flow prediction, outliers' detection, and demand prediction in urban transportation. That includes decision trees, Support Vector Machine or SVM, and random forests [5],[34].
2. Deep Learning (DL): Applied for heavy computing tasks including recognizing moving car images from traffic CCTV stream by using a deep learning model, CNNs, or for managing and maintaining the best route as per changing traffic patterns by using another type of neural nets, RNNs [18].
3. Generative AI: Rising as a powerful tool for representing urban samples, developing traffic patterns, and creating realistic synthetic data to train existing models [22],[26], [27], [49].
4. Reinforcement Learning (RL): Used for adaptive traffic signal controls to enhance transport along constantly changing roads and to help driverless cars navigate through urban environments [14].

RQ2: How do generative AI, machine learning, and deep learning contribute to advancements in urban mobility?

1. Generative AI:
 - Scenario Modeling: Allowed city planners to foresee the effects when adjusting concrete constructions.
 - Synthetic Data Generation: Overcome the problem of data deficit by creating more quality data for model learning particularly in the case of self-driving cars.
 - Advanced demand for the public transportation systems through changing timetables based on passengers' needs and requirements.
 - Management of electricity usage in mobility systems to optimize the energy efficiency of mobility systems [42].
 - Made great strides in real-time traffic monitoring through the use of video analytics on stream feeds and associating such data with Internet-of-Thing's sensors.
 - Assisted self-driving systems through the handling of information from various modalities, including vision, radar, and lidar [21], [23],[25].
2. Machine Learning:
 - Improved public transportation systems by forecasting passenger demand and adjusting schedules dynamically [11].
 - Enhanced energy efficiency in mobility systems by predicting and managing electricity usage [4].
3. Deep Learning:
 - Achieved significant advancements in real-time traffic monitoring by analyzing video feeds and integrating data from IoT sensors.
 - Supported autonomous driving systems by processing multimodal data (e.g., vision, radar, and lidar) [28], [29].

RQ3: What are the main applications of AI in urban mobility, such as traffic management, public transportation, autonomous vehicles, and environmental monitoring?

1. Traffic Management: About real-time traffic signal optimization, they had decreased the average travel time by 20-30%. Until Intelligent routing systems reduced traffic density in virtual heavily populated areas. Intelligent routing systems minimized congestion in densely populated areas [30], [31].
2. Public Transportation: About AI efficiency and effectiveness, AI-powered scheduling is said to have raised on-time performance by a quarter in pilots. Optimization of routes, and vehicle dispatch enhanced through the use of predictive models. Predictive models improved route planning and vehicle allocation [37-39].
3. Autonomous Vehicles: AI capability-based systems attained a twenty percent decrease in accident rates in urban environments for testing. Improved navigation and

safe driving are supported by the identification of real-time objects and decisions made by the developed algorithms [40].

4. Environmental Monitoring: AI applications for air pollution monitoring and reduction were effective, getting a 15% increase in the identification of air pollution hotspots. With the connection to the IoT networks, it offered detailed environmental information that supports the formulation of urban development strategies.

RQ4: How has AI been integrated with IoT and ICT to enhance urban mobility in smart cities?

1. IoT Integration: Traffic, environmental, and weather data was being gathered by smart sensors and devices. The real-life applications of IoT-connected AI systems included the offering of data-driven advice, for instance, intelligent traffic signal systems [46].
2. ICT Integration: Big Data solutions in the clouds handled large flows of urban mobility information for the following predictions. Mobile applications represented personal recommendations on the route and means of transportation using Artificial Intelligence [3],[49], [50].
3. Collaborative Systems: AI, IoT, and ICT improved communication between vehicles, pedestrians, and parts of the city such as smart crosswalks, and V2X. [1], [32]

RQ5: How do AI-driven urban mobility solutions differ across regions and urban contexts (e.g., developed vs. developing cities)?

1. Developed Cities: Concerned with the latest trends such as automatically operated automobiles and AI systems in transportation [11],[24]. Due to the significant IoT implementation, activities like data collection and analysis in a real-time environment became possible. There is a high use of AI mobility applications among urban dwellers.
2. Developing Cities: The most popular purpose of implementing AI was concerned with finding solutions that would create cost savings, for instance with further developments of existing forms of public transport and the avoidance of traffic congestion [13], [20], [38].

Relatively low development in this area as well as the lack of infrastructure weakened the implementation of better deep AI. The AI efforts were used for addressing particular problems, such as air contamination tracking or increasing safety on overcrowded streets.

RQ6: What are the ethical and social implications of AI-driven urban mobility solutions?

Ethical and Social Considerations in AI-Driven Urban Mobility The introduction of Artificial Intelligence for

urban transport creates multiple ethical and social issues which require specific solutions to develop fair and sustainable urban systems.

1. **Data privacy and Security:** AI-driven transportation system acquire their information through real-time monitoring of surveillance cameras and GPS tracking and smart sensors. Personal data safety along with protection against unauthorized access stands as a vital requirement. The security of user privacy demands both stringent rules and protection aimed at minimizing data security breaches. AI models face challenges regarding bias alongside
2. **Achieving bias practices in their functioning.** Different income levels create barriers that lead AI solutions to provide better services to resourceful areas while abandoning underserved populations. The public requires full visibility into AI decision systems to prevent discriminatory practices from occurring in transportation management networks
3. **Impact on Employment and Workforce**

Self-driving vehicles and AI-based public transport scheduling systems in transport systems pose job risk to drivers and transportation personnel. Social groups including both governments and industries need to launch training programs which will enable jobs transfer possibilities between existing and emerging mobility industry functions.

4. Public Trust and Acceptance

AI-driven mobility solutions will succeed only after the public develops trust in this particular technology. Approaches solving ethical matters about AI monitoring and administrative control systems must be developed to gain trust from urban citizens about their safety and security.

RQ7: How can AI-driven urban mobility solutions enhance disaster response and emergency management?

AI-driven urban mobility significantly improves the efficiency and effectiveness of disaster response by optimizing emergency transportation, predicting risks and quick resource allocation.

1. **Real-Time Traffic Management for Emergency Vehicles:** An intelligent traffic system with AI capabilities checks emergency vehicles first by manipulating signal control systems and finding the best routes for firefighters and ambulances together with police vehicles. First responders can use dynamic AI-based route optimization to find clear routes that let them reach their destination zones without delay.

2. Disaster Prediction and Risk Assessment: By

processing historical disaster data together with weather observation data alongside real-time sensor information AI models forecast future natural catastrophes like floods hurricanes and earthquakes. Through predictive analysis emergency authorities are able to initiate strategic actions which involve warning vulnerable areas followed by early emergency service deployments.

3. **Autonomous Drones and Robotics for Search and Rescue:** AI drones when deployed in disaster areas can automatically identify affected regions as well as detect survivor positions through their search capabilities and simultaneously distribute emergency rescue materials. Rescue operations receive aid from autonomous robots which use AI technology to complete tasks especially in challenging or dangerous access areas.
4. **Crowdsourced and IoT-Enabled Emergency Communication:** AI-powered systems analyze information collected from mobile devices in addition to social media and IoT sensors which generates immediate reports about damage areas and damaged infrastructure. Machine learning systems evaluate emergency telephonies together with social media alert systems to locate essential distress areas for emergency intervention.
5. **Smart Public Transportation for Evacuation Planning:** Disasters require the use of Artificial Intelligence to enhance public transportation operations by utilizing its ability to optimize stops of buses and trains and ride-sharing links to promote evacuation activities. Predictive models generate safe route evacuations which lead transportation resources to proper allocations.

This review examines the role of AI in enhancing urban mobility and advancing smart city initiatives. It highlights the diversity and adaptability of AI technologies, the specific challenges encountered in implementing AI solutions, and the disparities between developed and developing cities. By addressing these research questions, the study provides a comprehensive understanding of AI and its applications in the context of urban development.

Due to the various types of environmental and maintenance issues, hydropower generation in Sri Lanka has been going down seriously since the 1990s [24]. As a result, the Sri Lankan government has taken different steps. For example, electricity generation has transitioned to mixed hydro-thermal [20], [21].

As a result, the electricity generation of the country has transitioned to mixed hydro-thermal since 1998[20], [21]. In our work, annual power demand data from 2000 to 2022 is investigated. The observations were made from 2000 -2019 and are used for model fitting and 2020-2022 reserved ex-post testing

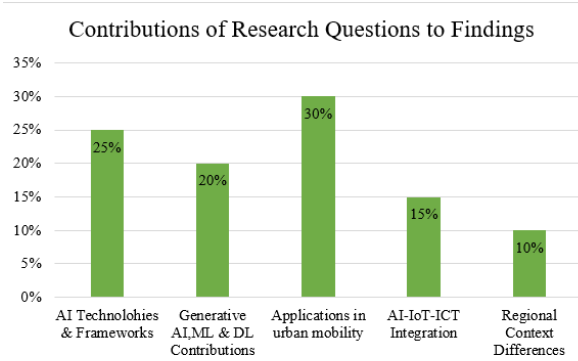


Figure 2: Contributions of Research Questions

IV. DISCUSSION

The discussion section recapitulates and analyzes the results of this Systematics Review, focusing on major conclusions, managerial implications, and research prospects for AI in smart cities and urban mobility.

A. AI's Transformative Potential in Urban Mobility

The current study analysis supports AI's essential function in transforming the landscape of travel in cities. In areas like real-time traffic monitoring, autonomous vehicle navigation, and self-driving cars, there are AI technologies including Machine learning, deep learning, and generative AI have proven significant enhancements in terms of efficacy, safety, and sustainability. Specifically:

- **Traffic Management:** Machine learning for adaptive traffic light timings, automatic route guidance, and other intelligent controls has been found highly effective in terms of minimizing congestion and average travel speeds even in large and growing urban centers in the developed and the developing world [19], [24], [35].
 - **Public Transportation:** The use of technology to forecast demands and schedule further avails the service has improved the commuting experience by making the services more reliable.
 - **Autonomous Vehicles:** Essential to deep learning technology self-driving cars can prevent accidents and improve driving in cities [33], [36].
 - **Environmental Monitoring:** AI has also enhanced the identification of areas of increased pollution, making cities employ the correct measures to deal with the issue [41], [43-45], [47], [48].
- These developments imply that AI is capable of handling challenges within urban mobility systems to increase scalability and flexibility and that is in the light of growing urbanization.

B. Integration of AI, IoT, and ICT

A combination of AI, IoT, and ICT has stepped up the capacity of mobility systems in urban centers. AI with IoT

and cloud ICT platforms for processing have facilitated the move of data as a medium in a feedback loop between human-constructed infrastructures.

- Smart IoT sensors with an AI interface have made adaptive traffic control systems better [15], [38], [46].
- Public transport: ICT platforms enabled the application of artificial intelligence through the development of applications that provide users and preferred routes as well as real-time updates [11], [19], [24].

However, the potential of integration of these two sectors remains partially unlocked in developing cities, the main reason being infrastructural limitations that exist.

Regional Disparities in AI Applications

A notable finding is the disparity in AI adoption between developed and developing cities:

- **Developed Cities:** These regions boast strong infrastructure, and they can develop new-generation AI applications such as self-driving cars and smart transportation [5], [9], [10].
- **Developing Cities:** These areas focus on the rationalization of tools including existing effective public transport, and solutions to sensitive problems including road safety and pollution.

They show the importance of having specific regional AI solutions so that everyone can benefit from innovations in urban mobility.

Challenges in Implementation

Despite its potential, several challenges hinder AI's broader adoption in urban mobility:

- **Data Privacy and Security:** What they fail to consider is that information needs to be processed and generated in real-time, a factor that increases the risk of hacking into the personal details of users that may be submitted to the sites [49].
- **Infrastructure and Cost Barriers:** Many emerging cities struggle to establish the necessary fundamentals of IoT empowered AI frameworks.
- **Algorithmic Bias:** Prejudice in the algorithms affects the probability of discriminated against people, especially when the

AI is designed for many people in urban areas. To make sure that AI-enabled mobility is fair and ethical, these challenges need to be met to support the reliable and fair applications of AI systems to solve urban mobility problems.

C. Research Gaps and Future Directions

This review identifies several areas for further research and innovation:

- **Generative AI in Scenario Planning:** While the potential of using generative AI for planning urban mobility is

high, its implementation has current limitations and should be researched more.

- **AI for Multimodal Mobility:** It is worth studying further the AI systems used in the context of multimodal transportation that is when individuals use bicycles, buses, and/or trains [17], [25], [40].
- **Sustainability Metrics:** Future studies should be directed toward developing a clearer measure of the environmental cost of AI-enabled mobility schemes.
- **Scalability for Developing Cities:** This is to call on more research in identifying affordable, large-scale AI applications that would best fit developing cities in terms of physical infrastructure [28], [29].

D. Practical Implications

The findings have practical implications for policymakers, urban planners, and AI developers:

- Therefore, it should be recommended that policymakers focus on the development of AI and IoT, especially in areas, which are characterized by high levels of urbanization.
- AI-based predictive solutions can be useful to urban planners for efficient resource management, and decision-making mechanisms.
- AI developers should emphasize the development of programmable and regional that contemplate the infrastructure availability problems also with the ethical aspects.

V. CONCLUSION

This review study aimed to observe the integration of AI-related concepts, technologies, and tools for urban development with the smart city concept in the modern era. The study comprised 50 research articles from reputed research databases. The study identified the usage of different technologies for Machine Learning, Deep Learning, IoT, Cloud Computing, and other computing approaches for smart city development while reducing vehicle traffic, protecting environmental aesthetics, automated vehicles, and other measures for urban development. This study has brought significant contributions to modern society by obtaining new knowledge to integrate AI-related technologies for the smart city concept.

REFERENCES

- [1] A. H. Alavi, P. Jiao, W. G. Buttlar, and N. Lajnef, "Internet of Things-enabled smart cities: State-of-the-art and future trends," *Measurement*, vol. 129, pp. 589–606, 2018, doi: 10.1016/j.measurement.2018.07.067.
- [2] V. Albino, U. Berardi, and R. M. Dangelico, "Smart cities: Definitions, dimensions, performance, and initiatives," *J. Urban Technol.*, vol. 22, no. 1, pp. 3–21, 2015, doi: 10.1080/10630732.2014.942092.
- [3] Z. Allam and Z. A. Dhunny, "On big data, artificial intelligence and smart cities," *Cities*, vol. 89, pp. 80–91, 2019, doi: 10.1016/j.cities.2019.01.032.
- [4] Z. Allam and A. Sharifi, "Research Structure and Trends of Smart Urban Mobility," *Smart Cities*, vol. 5, no. 2, pp. 539–561, 2022, doi: 10.3390/smartcities5020029.
- [5] S. S. Band et al., "When Smart Cities Get Smarter via Machine Learning: An In-Depth Literature Review," *IEEE Access*, vol. 10, pp. 60985–61015, 2022, doi: 10.1109/ACCESS.2022.3181718.
- [6] M. Batty, "Artificial intelligence and smart cities," *Environ. Plan. B: Urban Anal. City Sci.*, vol. 45, no. 1, pp. 3–6, 2018, doi: 10.1177/2399808317751169.
- [7] F. Cugurullo, "Urban Artificial Intelligence: From Automation to Autonomy in the Smart City," *Front. Sustain. Cities*, vol. 2, 2020, doi: 10.3389/frsc.2020.00038.
- [8] D. Diran, A. Fleur van Veenstra, T. Timan, P. Testa, and M. Kirova, "Adoption of innovative procurement procedures, entailing requirements for technical and ethically responsible AI," *Policy Dept. for Economic, Scientific and Quality of Life Policies*, European Parliament, 2021.
- [9] A. Kubik, "The Use of Artificial Intelligence in the Assessment of User Routes in Shared Mobility Systems in Smart Cities," *Smart Cities*, vol. 6, no. 4, pp. 1858–1878, 2023, doi: 10.3390/smartcities6040086.
- [10] A. Luusua, J. Ylipulli, M. Foth, and A. Aurigi, "Urban AI: understanding the emerging role of artificial intelligence in smart cities," *AI Soc.*, vol. 38, no. 3, pp. 1039–1044, 2023, doi: 10.1007/s00146-022-01537-5.
- [11] Z. Mahrez, E. Sabir, E. Badidi, W. Saad, and M. Sadik, "Smart Urban Mobility: When Mobility Systems Meet Smart Data," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6222–6239, 2022, doi: 10.1109/TITS.2021.3084907.
- [12] I. Mavlutova et al., "Urban Transportation Concept and Sustainable Urban Mobility in Smart Cities: A Review," *Energies*, vol. 16, no. 8, 2023, doi: 10.3390/en16083585.
- [13] P. A. M. S. A. Munhoz et al., "Smart mobility: The main drivers for increasing the intelligence of urban mobility," *Sustainability*, vol. 12, no. 24, pp. 1–25, 2020, doi: 10.3390/su122410675.
- [14] A. Nikitas, K. Michalakopoulou, E. T. Njoya, and D. Karampatzakis, "Artificial intelligence, transport and the smart city: Definitions and dimensions of a new mobility era," *Sustainability*, vol. 12, no. 7, pp. 1–19, 2020, doi: 10.3390/su12072789.
- [15] A. P. F. de Queiroz, D. S. Guimarães Júnior, A. M. Nascimento, and F. J. C. de Melo, "Overview of Urban Mobility in Smart Cities," *Res. Soc. Dev.*, vol. 10, no. 9, p. e18210917830, 2021, doi: 10.33448/rsd-v10i9.17830.

- [16] F. Shahzad, S. ur Rehman, A. Rehman Javed, Z. Jalil, and Y. bin Zikria, "Future Smart Cities: Requirements, Emerging Technologies, Applications, Challenges, and Future Aspects," *IEEE Commun. Surveys Tuts.*, vol. 1, n.d.
- [17] M. Shulajkovska et al., "Artificial Intelligence-Based Decision Support System for Sustainable Urban Mobility," *Electronics*, vol. 13, no. 18, 2024, doi: 10.3390/electronics13183655.
- [18] F. Andreescu, "The Current Scientific Stage of The Instruments and Methods Needed for an Efficient Traffic Management System Based on AI," *Inform. Econ.*, vol. 26, no. 1, pp. 46–56, 2022, doi: 10.24818/issn14531305/26.1.2022.05.
- [19] R. D. Chavhan and G. B. Sambare, "AI-Driven Traffic Management Systems In Smart Cities: A Review," *Educ. Adm. Theory Pract.*, vol. 2024, no. 5, pp. 105–116, 2024, doi: 10.53555/kuey.v30i5.2780.
- [20] N. V. Cuong and M. T. Aziz, "AI-Driven Vehicle Recognition for Enhanced Traffic Management: Implications and Strategies," n.d.
- [21] A. Degas et al., "A Survey on Artificial Intelligence (AI) and eXplainable AI in Air Traffic Management: Current Trends and Development with Future Research Trajectory," *Appl. Sci.*, vol. 12, no. 3, 2022, doi: 10.3390/app12031295.
- [22] S. Dikshit et al., "The Use of Artificial Intelligence to Optimize the Routing of Vehicles and Reduce Traffic Congestion in Urban Areas," *EAI Endorsed Trans. Energy Web*, vol. 10, pp. 1–13, 2023, doi: 10.4108/EW.4613.
- [23] A. Kadhodayi, M. Jabeli, H. Aghdam, and S. Mirbakhsh, "Artificial Intelligence-Based Real-Time Traffic Management," *J. Electr. Electron. Eng.*, vol. 2, no. 4, pp. 368–373, 2023.
- [24] O. I. Olayode, L. K. Tartibu, and M. O. Okwu, "Application of Artificial Intelligence in Traffic Control System of Non-autonomous Vehicles at Signalized Road Intersection," *Procedia CIRP*, vol. 91, pp. 194–200, 2020, doi: 10.1016/j.procir.2020.02.167.
- [25] A. A. Ouallane, A. Bahnasse, A. Bakali, and M. Talea, "Overview of Road Traffic Management Solutions based on IoT and AI," *Procedia Comput. Sci.*, vol. 198, pp. 518–523, 2021, doi: 10.1016/j.procs.2021.12.279.
- [26] M. R. Gopal Goriparthi, "Scalable AI Systems for Real-Time Traffic Prediction and Urban Mobility," *Int. J. Adv. Eng. Technol. Innov.*, vol. 1, 2021.
- [27] R. Šusteková, R. Knutelská, and M. Phd, "How Is the Artificial Intelligence Used in Applications for Traffic Management," n.d.
- [28] R. Abduljabbar, H. Dia, S. Liyanage, and S. A. Bagloee, "Applications of artificial intelligence in transport: An overview," *Sustainability*, vol. 11, no. 1, 2019, doi: 10.3390/su11010189.
- [29] A. Bennani, "AI-Based Approaches for Autonomous Vehicle Fleet Optimization and Management," *J. Artif. Intell. Res. Appl.*, vol. 3, n.d.
- [30] A. Falanga and A. Carteni, "Revolutionizing Mobility: Big Data Applications in Transport Planning," *WSEAS Trans. Environ. Dev.*, vol. 19, pp. 1421–1433, 2023, doi: 10.37394/232015.2023.19.129.
- [31] O. Igorevich Rozhdestvenskiy and E. Poornima, "Enabling Sustainable Urban Transportation with Predictive Analytics and IoT," *MATEC Web Conf.*, vol. 392, p. 01179, 2024, doi: 10.1051/mateconf/202439201179.
- [32] N. Japiassu, "AI-Powered Logistics and Mobility as a Service (MaaS): Driving the Future of Autonomous Vehicles and Smart Transportation," n.d., doi: 10.13140/RG.2.2.22609.75366.
- [33] A. Karpov, "AI-Driven Approaches for Autonomous Vehicle Fleet Coordination and Routing," *Aust. J. Mach. Learn. Res. Appl.*, vol. 3, no. 102, n.d.
- [34] G. Mancino, "The Role of Autonomous Vehicles and AI in Smart Transportation Systems: Enhancing Traffic Management, Route Optimization, and Predictive Maintenance." [Online]. Available: <https://doi.org/10.13140/RG.2.2.23868.04483>
- [35] R. Mohammed, "Artificial Intelligence-Driven Robotics for Autonomous Vehicle Navigation and Safety." [Online]. Available: <https://nexusaireview.com/>
- [36] M. Nadeem, "AI in Autonomous Vehicles: State-of-the-Art and Future Directions," *Int. J. Adv. Eng. Technol. Innov.*, vol. 1, 2024.
- [37] B. N. Silva et al., "Urban planning and smart city decision management empowered by real-time data processing using big data analytics," *Sensors*, vol. 18, no. 9, 2018, doi: 10.3390/s18092994
- [38] J. Singh, "AI-Driven Path Planning in Autonomous Vehicles: Algorithms for Safe and Efficient Navigation in Dynamic Environments," *J. AI-Assist. Sci. Discov.*, vol. 4, n.d.
- [39] J. Singh, "Autonomous Vehicle Swarm Robotics: Real-Time Coordination Using AI for Urban Traffic and Fleet Management," *J. AI-Assist. Sci. Discov.*, vol. 3, no. 2, n.d.
- [40] J. Singh, "Autonomous Vehicles and Smart Cities: Integrating AI to Improve Traffic Flow, Parking, and Environmental Impact," *J. AI-Assist. Sci. Discov.*, vol. 4, n.d.
- [41] D. Ushakov, E. Dudukalov, L. Shmatko, and K. Shatila, "Artificial Intelligence as a factor of public transportation system development," *Transp. Res. Procedia*, vol. 63, pp. 2401–2408, 2022, doi: 10.1016/j.trpro.2022.04.001.

- 10.1016/j.trpro.2022.06.276
- [42] S. Andreescu and O. A. Sadik, "Trends and challenges in biochemical sensors for clinical and environmental monitoring," *Pure Appl. Chem.*, vol. 76, no. 4, 2004.
- [43] L. Capelli, S. Sironi, and R. del Rosso, "Electronic noses for environmental monitoring applications," *Sensors*, vol. 14, no. 11, pp. 19979–20007, 2014. doi: 10.3390/s141119979
- [44] C. K. Ho, A. Robinson, D. R. Miller, and M. Davis, "Overview of Sensors and Needs for Environmental Monitoring," *Sensors*, vol. 5, pp. 4–37, 2005. [Online]. Available: <http://www.mdpi.org/sensors>
- [45] O. Igorevich Rozhdestvenskiy and E. Poornima, "Enabling Sustainable Urban Transportation with Predictive Analytics and IoT," *MATEC Web Conf.*, vol. 392, p. 01179, 2024. doi: 10.1051/mateconf/202439201179
- [46] F. Long, A. Zhu, and H. Shi, "Recent advances in optical biosensors for environmental monitoring and early warning," *Sensors*, vol. 13, no. 10, pp. 13928–13948, 2013. doi: 10.3390/s131013928
- [47] H. Messer, A. Zinevich, and P. Alpert, "Environmental monitoring by wireless communication networks," *Science*, vol. 312, no. 5774, p. 713, 2006. doi: 10.1126/science.1120034
- [48] L. Shu, H. H. Chen, T. Hara, D. J. Deng, and L. Wang, "Guest editorial," *J. Commun.*, vol. 6, no. 2, pp. 125–127, 2011. doi: 10.4304/jcm.6.2.125-127
- [49] B. N. Silva et al., "Urban planning and smart city decision management empowered by real-time data processing using big data analytics," *Sensors*, vol. 18, no. 9, 2018. doi: 10.3390/s18092994
- [50] R. Abduljabbar, H. Dia, S. Liyanage, and S. A. Bagloee, "Applications of artificial intelligence in transport: An overview," *Sustainability*, vol. 11, no. 1, p. 189, 2019. doi: 10.3390/su11010189
- [51] A. Addas, M. Tahir, and N. Ismat, "Enhancing precision of crop farming towards smart cities: An application of artificial intelligence," *Sustainability*, vol. 16, no. 1, p. 123, 2023. doi: 10.3390/su16010123
- [52] M. Al-Hader and A. Rodzi, "The smart city infrastructure development & monitoring," *Theor. Empir. Res. Urban Manag.*, vol. 4, no. 2, pp. 87–94, 2009.
- [53] V. Albino, U. Berardi, and R. M. Dangelico, "Smart cities: Definitions, dimensions, performance, and initiatives," *J. Urban Technol.*, vol. 22, no. 1, pp. 3–21, 2015. doi: 10.1080/10630732.2014.942092
- [54] E. Almeshaie, A. Al Habaibeh, and B. Shakhmak, "Rapid evaluation of micro-scale photovoltaic solar energy systems using empirical methods combined with deep learning neural networks," *J. Clean. Prod.*, vol. 244, p. 118788, 2019. doi: 10.1016/j.jclepro.2019.118788
- [55] S. U. Amin et al., "Cognitive smart healthcare for pathology detection and monitoring," *IEEE Access*, pp. 10745–10753, 2019. doi: 10.1109/ACCESS.2019.2891390
- [56] B. Anthony Jnr, "A case-based reasoning recommender system for sustainable smart city development," *AI Soc.*, vol. 36, pp. 1–25, 2020. doi: 10.1007/s00146-020-00984-2
- [57] M. S. Asad et al., "Mobility prediction-based optimisation and encryption of passenger traffic-flows using machine learning," *Sensors*, vol. 20, no. 9, p. 2629, 2020. doi: 10.3390/s20092629
- [58] A. Awan, M. Alnour, A. Jahanger, and J. C. Onwe, "Do technological innovation and urbanization mitigate carbon dioxide emissions from the transport sector?," *Technol. Soc.*, vol. 71, p. 102128, 2022. doi: 10.1016/j.techsoc.2022.102128
- [59] L. Bartolucci et al., "PV assisted electric vehicle charging station considering the integration of stationary first- or second-life battery storage," *J. Clean. Prod.*, vol. 383, p. 135426, 2023. doi: 10.1016/j.jclepro.2022.135426
- [60] S. Benedict, "Shared mobility intelligence using permissioned blockchains for smart cities," *New Gener. Comput.*, vol. 40, no. 4, pp. 1009–1027, 2022. doi: 10.1007/s00354-021-00147-x
- [61] A. R. Ben, M. Sehl, and C. Jacqueline, "Barriers to artificial intelligence adoption in smart cities: A systematic literature review and research agenda," *Gov. Inf. Q.*, vol. 40, no. 3, p. 101761, 2023. doi: 10.1016/j.giq.2023.101761
- [62] L. Butler, T. Yigitcanlar, and A. Paz, "Barriers and risks of mobility-as-a-service (MaaS) adoption in cities: A systematic review of the literature," *Cities*, vol. 109, p. 103036, 2021. doi: 10.1016/j.cities.2020.103036
- [63] J. D. J. Camacho et al., "Leveraging artificial intelligence to bolster the energy sector in smart cities: A literature review," *Energies*, vol. 17, no. 2, p. 345, 2024. doi: 10.3390/en17020345
- [64] J. Cao et al., "Urban noise recognition with convolutional neural network," *Multimed. Tools Appl.*, vol. 78, no. 20, pp. 29021–29041, 2019. doi: 10.1007/s11042-019-07793-4
- [65] A. Caragliu, C. Del Bo, and P. Nijkamp, "Smart cities in Europe," *J. Urban Technol.*, vol. 18, no. 2, pp. 65–82, 2011. doi: 10.1080/10630732.2011.601117
- [66] D. Carracedo and H. Mostofi, "Electric cargo bikes in urban areas: A new mobility option for private transportation," *Transp. Res. Interdiscip. Perspect.*, vol. 16, p. 100705, 2022. doi: 10.1016/j.trip.2022.100705

- [67] M. Castelli, R. Sormani, L. Trujillo, and A. Popovič, "Predicting per capita violent crimes in urban areas: An artificial intelligence approach," *J. Ambient Intell. Humaniz. Comput.*, vol. 8, no. 1, pp. 29–36, 2017. doi: 10.1007/s12652-016-0366-8
- [68] M. Chen et al., "Cognitive-LPWAN: Towards intelligent wireless services in hybrid low power wide area networks," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 2, pp. 402–417, 2019. doi: 10.1109/TGCN.2019.2897075
- [69] W. Chen, Y. Chen, X. Chen, and Z. Zheng, "Toward secure data sharing for the IoV: A quality-driven incentive mechanism with on-chain and off-chain guarantees," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1625–1640, 2020. doi: 10.1109/JIOT.2019.2946611
- [70] J. Culita et al., "A hybrid approach for urban traffic prediction and control in smart cities," *Sensors*, vol. 20, no. 24, p. 7209, 2020. doi: 10.3390/s20247209
- [71] S. Danuta et al., "Artificial intelligence in the smart city — A literature review," *Eng. Manag. Prod. Serv.*, vol. 15, no. 4, pp. 53–75, 2023. doi: 10.2478/emj-2023-0027
- [72] S. B. Elias et al., "Environmentally sustainable smart cities and their converging AI, IoT, and big data technologies and solutions: An integrated approach to an extensive literature review," *Energy Inform.*, vol. 6, no. 1, p. 9, 2023. doi: 10.1186/s42162-023-00253-0
- [73] S. B. Elias et al., "Smarter eco-cities and their leading-edge artificial intelligence of things solutions for environmental sustainability: A comprehensive systematic review," *Environ. Sci. Ecotechnol.*, vol. 19, 2024.
- [74] A. B. Haque, A. Muniat, P. R. Ullah and S. Mushsharat, "Automated approach towards smart healthcare with blockchain and smart contracts," in *Proc. 2021 Int. Conf. Comput., Commun. Intell. Syst. (ICCCIS)*, pp. 250–255, 2021, doi: 10.1109/ICCCIS51004.2021.9397187.
- [75] L. Kong, "A study on the AI-based online triage model for hospitals in sustainable smart city," *Future Gener. Comput. Syst.*, vol. 125, pp. 59–70, 2021, doi: 10.1016/j.future.2021.06.035.
- [76] A. Kubik, "The use of artificial intelligence in the assessment of user routes in shared mobility systems in smart cities," *Smart Cities*, vol. 6, no. 4, pp. 1858–1878, 2023, doi: 10.3390/smartcities6040086.
- [77] D. Liang and Y. Yunhong, "Frontiers of policy and governance research in a smart city and artificial intelligence: An advanced review based on natural language processing," *Front. Sustain. Cities*, vol. 5, p. 123456, 2023, doi: 10.3389/frsc.2023.123456.
- [78] J. D. Martin, "Frankenstein urbanism: Eco, smart and autonomous cities, artificial intelligence and the end of the city," *J. Urban Technol.*, vol. 30, no. 4, pp. 113–114, 2023, doi: 10.1080/10630732.2023.2235123.
- [79] P. Persaud, A. S. Varde and S. Robila, "Enhancing autonomous vehicles with commonsense: Smart mobility in smart cities," in *Proc. 2017 IEEE 29th Int. Conf. Tools Artif. Intell. (ICTAI)*, pp. 1008–1013, 2017, doi: 10.1109/ICTAI.2017.00155.
- [80] R. Polishetty, M. Roopaei and P. Rad, "A next-generation secure cloud-based deep learning license plate recognition for smart cities," in *Proc. 2016 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, pp. 286–293, 2016, doi: 10.1109/ICMLA.2016.0055.
- [81] F. Scorza and G. Fortunato, "Cyclable cities: Building feasible scenario through urban space morphology assessment," *J. Urban Plan. Dev.*, vol. 147, no. 4, p. 05021034, 2021, doi: 10.1061/(ASCE)UP.1943-5444.0000713.
- [82] S. A. Shaheen, S. Guzman and H. Zhang, "Bikesharing in Europe, the Americas, and Asia: Past, present, and future," *Transp. Res. Rec.*, vol. 2143, no. 1, pp. 159–167, 2010, doi: 10.3141/2143-20.
- [83] J. Siderska and K. S. Jadaan, "Cloud manufacturing: A service-oriented manufacturing paradigm. A review paper," *Eng. Manag. Prod. Serv.*, vol. 10, no. 1, pp. 22–31, 2018, doi: 10.1515/emj-2018-0002.
- [84] S. K. Singh, S. Rathore and J. H. Park, "BlockIoTIntelligence: A blockchain-enabled intelligent IoT architecture with artificial intelligence," *Future Gener. Comput. Syst.*, vol. 110, pp. 721–743, 2020, doi: 10.1016/j.future.2019.09.002.
- [85] A. Solanas et al., "Smart health: A context-aware health paradigm within smart cities," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 74–81, 2014, doi: 10.1109/MCOM.2014.6871673.
- [86] D. Szpilko, J. Szydło and J. Winkowska, "Social participation of city inhabitants versus their future orientation: Evidence from Poland," *WSEAS Trans. Bus. Econ.*, vol. 17, pp. 692–702, 2020, doi: 10.37394/23207.2020.17.66.
- [87] K. Szum, "IoT-based smart cities: A bibliometric analysis and literature review," *Eng. Manag. Prod. Serv.*, vol. 13, no. 2, pp. 115–136, 2021, doi: 10.2478/emj-2021-0015.
- [88] S. Toglaw, M. Aloqaily and A. A. Alkheir, "Connected, autonomous and electric vehicles: The optimum value for a successful business model," in *Proc. 2018 5th Int. Conf. Internet Things: Syst., Manag. Security (IoTSMS)*, pp. 303–308, 2018, doi: 10.1109/IoTSMS.2018.8554685.
- [89] E. J. Tomaszewska and A. Florea, "Urban smart mobility in the scientific literature—Bibliometric analysis," *Eng. Manag. Prod. Serv.*, vol. 10, no. 2, pp. 41–56, 2018, doi: 10.2478/emj-2018-0010.
- [90] J. Yan, G. Zhaoquan, D. Lei et al., "Artificial

intelligence enabled cyber security defense for smart cities: A novel attack detection framework based on the MDATA model," *Knowl.-Based Syst.*, vol. 276, p. 110753, 2023, doi: 10.1016/j.knosys.2023.110753.

- [91] T. Yigitcanlar, N. Kankanamge and K. Vella, "How are smart city concepts and technologies perceived and utilized? A systematic geo-twitter analysis of smart cities in Australia," *J. Urban Technol.*, vol. 28, no. 1–2, pp. 135–154, 2021, doi: 10.1080/10630732.2020.1753483.
- [92] T. Yigitcanlar, R. Mehmood and J. M. Corchado, "Green and smart: Towards sustainable urban futures with artificial intelligence," *Sustainability*, vol. 13, no. 4, p. 1923, 2021, doi: 10.3390/su13041923.
- [93] H. Zheng, X. Chen and X. M. Chen, "How does on demand ridesplitting influence vehicle use and purchase willingness? A case study in Hangzhou, China," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 3, pp. 143–157, 2019, doi: 10.1109/MITS.2019.2919503.
- [94] Y. Zhenjun, J. Ling, H. Xiaoli et al., "Intelligent urbanism with artificial intelligence in shaping tomorrow's smart cities: Current developments, trends, and future directions," *J. Cloud Comput.*, vol. 12, no. 1, p. 65, 2023, doi: 10.1186/s13677-023-00427-3.

ACKNOWLEDGMENT

I am deeply grateful for the support and guidance I received throughout this SLR. I sincerely thank all those who contributed to the successful completion of this work. My heartfelt appreciation goes to the academic staff of the Department, the institutions that provided access to research resources, and all who shared valuable insights and constructive feedback. I also extend my gratitude to my family and friends for their continuous encouragement and motivation during this journey.

AUTHOR BIOGRAPHIES

K. Luxshi

Luxshi is an undergraduate student at Sabaragamuwa University of Sri Lanka with a strong interest in Machine Learning and Artificial Intelligence. Her academic journey is driven by a passion for innovation and a deep curiosity about emerging technologies. She has gained hands-on experience in both programming and research and is committed to exploring the practical applications of AI to address real-world challenges. She is keen to collaborate on impactful research and continues to expand her knowledge in the field.

Professor R.M.K.T. Rathnayake

Professor R.M.K.T. Rathnayake is a Full Professor in the Department of Physical Science and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka. Currently, he is the Dean of

the Faculty of Applied Sciences, at Sabaragamuwa University of Sri Lanka. He has completed Doctoral and Master's degrees from the Wuhan University of Technology in China. Further, he has completed a Master's degree from the University of Sri Jayewardenepura. His first degree was completed from the University of Ruhuna with a first-class honours degree

Dr. D.M.K.N. Seneviratna

Dr. D.M.K.N. Seneviratna holds a BSc (Special), MSc, and PhD, and currently serves as a Senior Lecturer at the Faculty of Engineering, University of Ruhuna, Sri Lanka. Her academic and research expertise spans across engineering and computing disciplines, with a strong commitment to advancing knowledge through teaching, research, and academic mentorship.

Mr. W.M.C.J.T. Kithulwatta

Mr. W.M.C.J.T. Kithulwatta is a Lecturer in the Department of Information and Communication Technology, Faculty of Technological Studies, Uva Wellassa University of Sri Lanka. He has completed his Master of Philosophy degree at Sabaragamuwa University of Sri Lanka in the area of Computer Science and he has completed his first degree with a BSc (Hons) in Software Engineering from the University of Kelaniya with first- class honours.

Innovative ECG Classification Approach Utilizing a Transfer Learning-Driven Ensemble Architecture

HMLS Kumari^{1#}

¹Computer Center, Faculty of Engineering, University of Peradeniya, Sri Lanka

[#]ihinisangeetha99@gmail.com

ABSTRACT An electrocardiogram (ECG/EKG) is a vital methodology that is used for the diagnosis and monitoring of heart diseases by recording the electrical activity of the heart. However, manual analysis of ECGs shows limitations such as noise sensitivity, visual interpretation constraints and data imbalance. The proposed study a deep learning ensemble model combining DenseNet121, InceptionV3, and ResNet50 are implement to classify ECG images to improve diagnostic accuracy. The model is trained on two datasets: the National Heart Foundation 2023 ECG dataset and the ECG Dataset for Heart Condition Classification, focusing the main cardiac conditions such as abnormal heartbeat, myocardial infarction. The preprocessing techniques include background removal of ECG signal images, grayscale conversion, and data augmentation to enhance image quality and overfitting reduction. Stratified 5-Fold cross-validation was employed to demonstrate the generalization abilities of the proposed models. Early stopping and performance plots demonstrated that proposed model is not overfitting and two proposed models show consistent accuracy which suggests the model is not biased toward a specific dataset. While the ensemble models, as demonstrated in this study, produce better results than single models. The proposed study demonstrates validation accuracies of 98.62% and 96.75% for the National Heart Foundation 2023 dataset and the ECG dataset for heart condition classification, respectively, using 5-fold stratified cross-validation. There are still some limitations, such as the proposed ensemble models not being evaluated using Explainable AI, which reduces clinical trust. Additionally, small datasets can limit the model's generalizability. Therefore, this study demonstrates the potential of deep ensemble models with advanced preprocessing for ECG classification, but it also highlights the importance of greater transparency, better dataset diversity, and real-world validation in future research studies.

INDEX TERMS Data Augmentation, ECG (electrocardiogram), Ensemble Architectures, Preprocessing techniques, Transfer learning

I. INTRODUCTION

An electrocardiogram involves placing temporary electrodes on chest and limbs to record heart's electrical signals that regulate its beats. This information is analysed by a computer and is shown as a wave pattern the healthcare professional can further analyse for diagnostic insight. Figure 1 shows the components of the heart that are involved in an ECG. The internal pacemaker is a natural system in the heart that controls its rhythm and rate. The sinoatrial (SA) node, or what is commonly called the internal pacemaker of the heart. The SA node initiates the heartbeat by firing an electrical impulse. An ECG detects this electrical signal and tracks its path as heart contracts and then relaxes with each heartbeat. The electrical activity of the heart produces three distinct waves, each corresponding to specific actions within the heart. The first wave, the "P wave," is produced by the upper chambers of the heart, known as the atria, where the heartbeat originates. The "QRS complex" is the wave produced by the contraction of the lower chambers, the ventricles. Finally, the third wave, called the "T wave," represents the recovery or resting phase of the heart following a beat [1]. Figure 2 shows an ECG wave for a

normal person, and Figure 3 depicts the general representation of the wave in a 1-heartbeat process. Classification of abnormalities in the ECG signals is not an easy task in any way because several obstacles lay the process of classification. Firstly, analysis or identification directly on the paper recordings of the ECG is a huge problem.

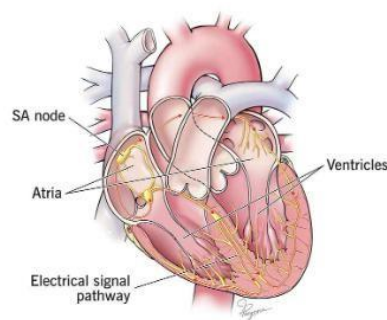


Figure 1. The diagram of heart with its components that helps to get ECG graphs [1]



Figure 2. The normal person ECG wave [1]



Figure 3. The normal person ECG wave [1]

In manual diagnosis, medical professionals often find it hard to interpret ECG signals using simple visual examination because waveforms can be complex and have subtle differences that can be missed. To help with that, deep learning models based on advanced image and signal processing have been developed. These models are considered to provide more accurate and reliable decisions than traditional human methods. However, there are some drawbacks in these models, such as data noise, unbalanced case types, and differences in patient conditions. Machine learning methods have shown improved results, but their accuracy is still limited, especially in handling changing situations based on patient conditions.

Recent research further enhances deep learning, which can automatically identify important features in ECG images and signals. Techniques like CNNs and RNNs have worked well, but they do not demonstrate high accuracy when used alone. Therefore, ensemble models where several models are combined to form a stronger system can be introduced, though this approach is not yet widely used. Data augmentation, which involves generating more varied training samples, is also useful for preventing overfitting and improving model performance, especially when the dataset is small. Stratified 5-fold cross-validation was used to evaluate the reliability and generalizability of the predicted decisions from the proposed models. Even as the proposed study claims novelty through a method of ensemble learning, the contribution is incremental, rather than revolutionary. The methodology employs an ensemble of well-known pre-trained CNN architectures such as DenseNet121, InceptionV3, and ResNet50, combined by a simple average ensemble approach with deep layer fine-tuning, in addition to the addition of dense layers, dropout, and batch normalization. Technically valid, the methodology in effect reproduces existing architectures without introducing a fundamentally new learning algorithm or architectural innovation. But while the ensemble approach may not constitute a conceptual innovation, it offers a solution to a key practical issue in medical image

classification, enhancing robustness and generalizability via suitably tuned ensembles of pre-existing CNN models. The addition of stratified 5-fold cross-validation, dropout regularization, and equable performance across folds further reinforces the clinical credibility of the findings. Furthermore, parallel application of multiple architectures allows complementary feature extraction, offering an effective but practical means of improving ECG classification performance particularly in situations where resources are constrained or in real-time diagnostic contexts. Nevertheless, the study can be strengthened by exploring more innovative alternatives, i.e., Transformer-based methods, or comparison to different ensemble techniques and advanced fusion mechanisms for better basis to justify its novelty claims.

II. RELATED WORKS

Mohammed et al. [2] proposed a CNN that was especially designed for the classification of ECG signals. The study focuses on the classification of five types of ECG arrhythmic signals using the MIT-BIH Arrhythmia Dataset from PhysioNet. ECG preprocessing techniques such as denoising, ORS peak detection, and heart beat segmentation were performed. 1-D CNN ECG heart beat classifier was performed and achieved a result of an accuracy of 95.2% achieved. This study outlines how deep learning models depend on the prior quality of input data to function optimally through preprocessing steps in the medical imaging field. Using a single dataset makes it difficult to generalize the model's application. Without employing techniques like regularization, data modification, or outside validation, the chances of overfitting rise, which adversely affects the model's usefulness in actual clinical environments. The ensemble model, which was proposed by Essa et al. [3], combines a deep learning-based multi-model of a CNN with an LSTM network. As a result, can conclude that multi models have shown better performance than single models.

The other known system adheres to the hybrid classification technique of ECG images obtained from the "ECG Images dataset of Cardiac Patients" done by tariq et al., [4]. Those included ECG images with 12-lead ECG classified under abnormal heartbeat, Myocardial Infraction (MI), prior MI history, and normal ECG. With the integration of IoT-based data acquisition with a CNN-based classifier integrated with an attention module, excellent accuracy of 98.39% was realized, depicting the potential of this proposed system in identifying cardiac disorder diagnosis. Acharya et al. [5] proposed a 13-layer deep fully convolutional neural network for classification. Although the proposed methodology performed very efficiently, it involves complex computational time and is bound to have limitations in practice; the results would vary based on expertise. Similar to ANNs, the performance of a CNN also depends upon structure,

weights, and preferences of earlier layers. To improve the efficiency and reduce overfitting, pooling was used that reduced the size of output generated from convolutional layers and consequently reduced computational demands. The final accuracy is 88.67%, the specificity 90.00%, and a sensitivity of 95.00% respectively. Fatema et al. [6] proposed a deep learning approach for cardiovascular diseases classification using the enhanced paper-based electrocardiogram images. The work focused on the optimization of precision with a minimal time complexity to classify heart disorders into five classes. Different image preprocessing techniques were used to improve the quality and eliminate artifacts of the images before training. Besides, model components and hyperparameter optimization can be done via an ablation study, thereby improving the performances even more. Indeed, the InRes-106 model recorded an impressive accuracy of 98.34%, higher than any individual models (InceptionV3: 90.56%, ResNet50: 89.63%, DenseNet201: 88.94%, VGG19: 87.87%, MobileNetV2: 80.56%).

Deep Learning framework for cardio vascular disease prediction using ECG Images done by Muthu Meena et al. [7] proposed a study on the ECG image dataset to predict four primary cardiac conditions. These include an abnormal heartbeat, myocardial infarction, history of myocardial infarction, and normal cases using deep learning techniques. A hybrid model combining Inception-V3 and VGG-19 is proposed for cardiovascular disease prediction. Inception-V3 has excellent feature extraction capability with much efficiency, while VGG-19 extracts spatial information in minute detail from ECG images. This hybrid model performs the task with better results by considering the complementary strengths of these architectures. The proposed model significantly outperforms previous studies in terms of an accuracy of 96.0%, precision of 90.90%, recall of 100%, and F1-score of 95.23%.

III. METHODOLOGY

To classify ECG to detect heart diseases, Averaging Ensemble of DenseNet121, InceptionV3, and ResNet50 models is utilized. Before that, several preprocessing methods has been applied to those ECG images and then classification done.

A. Data Acquisition

1. National Heart Foundation 2023 ECG dataset

National heart foundation 2023 ECG dataset [8] is a free open dataset available on the Kaggle website. This is an ECG image dataset from the National heart foundation of Bangladesh, and it presents a collection of ECG images with various categories related to cardiac health. These images represent the heart's electrical activity versus time and convey critical information about the health status of

the heart. Basically, there are four classes of subjects in the dataset: abnormal heartbeats, myocardial infarction, normal, and previous myocardial infarction.

Abnormal Heartbeat Patients: This category contains ECG images taken from patients who have irregular heartbeats or arrhythmias. Examples include Atrial fibrillation, bradycardia, tachycardia, and ectopic beats and other abnormal rhythm conditions [9]. These image types help analyse and diagnose various heart-related ailments. **Myocardial Infarction class** represents ECG images that are grouped into this category and are from patients diagnosed with myocardial infarction, commonly referred to as a heart attack. The recordings often reveal distinct changes in the waveform, such as ST-segment elevation or depression, T-wave inversion, or pathological Q-waves, signifying myocardial damage [10]. **Normal Individuals** represent group comprises ECG readings from individuals without detectable cardiac abnormalities. These recordings serve as standard reference patterns, enabling comparison with abnormal ECGs to identify deviations from normal cardiac activity [8]. The category **Myocardial Infarction class history** includes ECG images of patients with a previous history of myocardial infarction. Images depict changes that may be permanent relative to prior myocardial damage or transient variations seen at follow-up studies or during continuous cardiac monitoring [8]. The total amount of ECG images is 2898, and further details of classes in the above dataset are shown in Table 1. As shown in table1 above dataset is balanced.

Table 1. The brief details of number of ECG mages in each class of dataset National heart foundation ECG 2023 dataset

Class	No. ECG images
Abnormal Heartbeat	814
Myocardial Infraction	716
Normal Person	852
Patients that have history of Myocardial Infraction	516

2. ECG dataset for Heart condition classification

The proposed study is not biased to a single dataset. Hence, another dataset considered here is the ECG Dataset for Heart Condition Classification [11]. This is freely available in the Kaggle website. The categories included in this dataset are three classes of ECG signals: normal signals, abnormal rhythms, and cardiac signals arising due to some diseases. It includes records that are labelled and collected from healthy individuals and patients with different heart conditions. The data will be used for machine learning-based applications that can allow one to continuously monitor a person's health and predict cardiac diseases in real time. The dataset also maintains an appropriate balance of good-quality ECG images across classes. Therefore, the dataset offers access to powerful tools with

AI-enabled diagnostic support in healthcare. There are 707 images in the combined dataset. Table 2 demonstrates the brief details of number of ECG mages in each class of ECG for Heart condition classification dataset

Table 2. Number of ECG images in each class of the dataset ECG for Heart condition classification

Class	Number of ECG images
Abnormal heart beat	241
History of Myocardial Infraction	171
Normal Person	295

3. ECG Signal Preprocessing

The original image of national heart foundation 2023 ECG dataset is shown in Figure 4 and represents additional background in the ECG image. Therefore, the proposed approach followed the background removal preprocessing step of the ECG image so that only the ECG signal would be considered. Figure 5 presents the pre-processed ECG image.

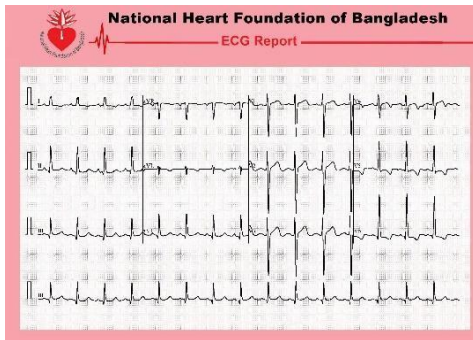


Figure 4. The ECG image of Abnormal heart beat class of National Heart Foundation 2023 ECG dataset.

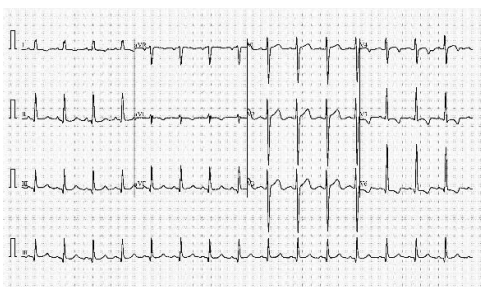


Figure 5. The cropped ECG image

ECG image extraction by removing the background was done only for the National Heart Foundation 2023 ECG dataset.

The preprocessing technique is applied on the ECG image by changing into grayscale[12], normalizing for increasing the contrast of the ECG image, after which Gaussian Blur[13] was applied in reducing noise using the Gaussian kernel size of 5, Otsu's thresholding [14] used to change the

blurred image into a binary image, resize the image to 256X 256. Use the above preprocessing techniques to make 1 image from the original dataset and add those images to the original datasets. As a result, each image makes new 5 images according to the preprocessing steps. Figure 6 shows the samples of pre-processed steps used for all classes of dataset National heart foundation ECG 2023 dataset.

After splitting new dataset into train 70%, test 20% and validation 10% by total dataset and train the proposed Ensemble CNN model. Few data augmentation steps are used to reduce overfitting before training.

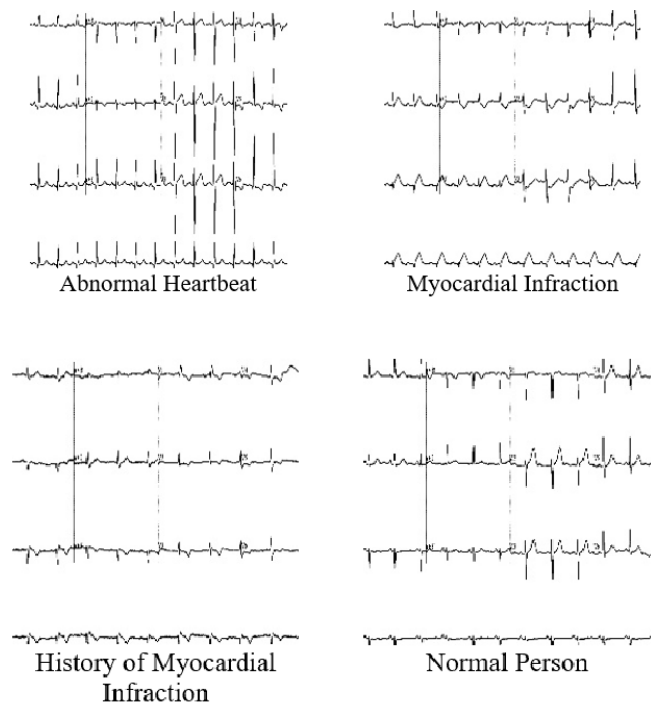


Figure 6. The resulted ECG images after preprocessing techniques.

4. Ensemble Model for CNN Transfer Learning

As shown in Figure 7, The proposed study proposed an ensemble model of DenseNet121, InceptionV3 and Resnet50.

The top layer of each transfer learning model is removed (include_top=False), and ImageNet weights are used. To retain the pre-trained features of the models, 80% of the layers are frozen. A global average pooling layer, a dense layer with 1024 neurons using the ReLU activation function are added to each model's output. These steps are followed individually for all three transfer learning models. The outputs from each model are averaged using the Average() layer. The resulting fused model is passed through a Dense layer (1024 neurons with ReLU activation), followed by a Dropout layer (0.4) for regularization and Batch

Normalization for training stability. The final layer is an output layer consisting of a Dense layer, with the number of units depending on the number of classes in the classification problem.

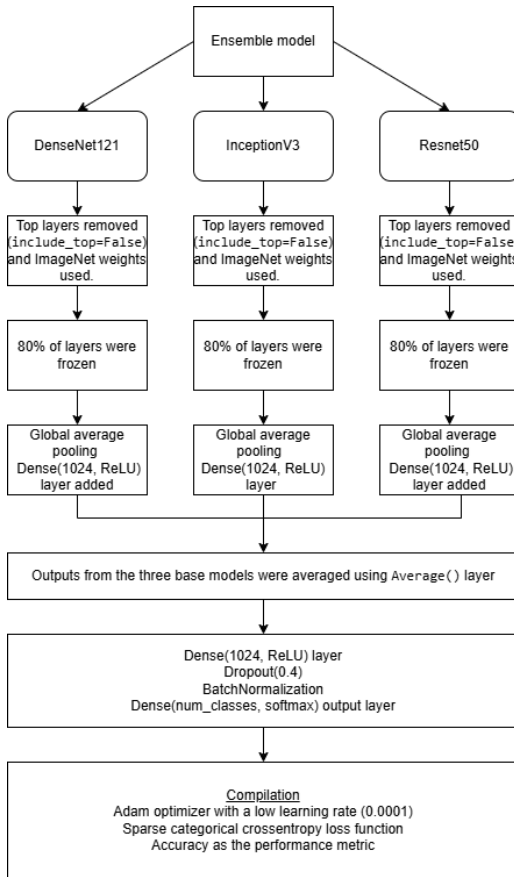


Figure 7. The proposed ensemble model

Using an Adam optimizer with a learning rate of 0.0001, the model was trained as per the schedule outlined in the provided performance evaluation method employing sparse categorical cross entropy loss. For better class balance within folds, given the small size of the datasets and to ensure robust evaluation with various data splits, a 5-fold Stratified K-Fold Cross-Validation was implemented. During both the training and evaluation phases, each iteration of the folds was trained as a fresh model to reduce overfitting, incorporating early stopping and learning rate reduction callbacks to optimize performance.

Model performance was evaluated on the training and validating sets, and the training accuracy, validation accuracy, precision, recall, and F1 score for validation were calculated for each fold.

IV. RESULTS AND DISCUSSION

Using the ensemble model of the transfer learning approach, suboptimal accuracy was achieved. This performance could be

mainly due to the limited size of the dataset. Small datasets often result in overfitting of the model on the training data while failing on the unseen test data [15]. To overcome this, early stopping and 5-Fold Stratified cross validation are used while training the model

The proposed model is based on an ensemble model of Densenet121, Inception V3, and Resnet50 for the above two datasets. Other than that, we used to train individual models like Densenet121, InceptionV3, and Resnet50 for the above dataset separately. But finally, we got high accuracy for the ensemble model other than an individual model of Densenet121, Resnet 50, and InceptionV3. This observation can be seen in both datasets used in proposed work. The proposed model was trained by taking the 50 epochs, batch size 32 with learning rate 0.0001, dense layer 1024 along with the 'relu' activation layer, and the dropout 0.4 National Heart Foundation 2023 ECG dataset with 4 classes and the ECG dataset to classify Heart conditions with 3 classes, respectively.

Table 3 demonstrates the metrics of the best performance achieved at validation in 5-fold stratified cross validation which are validation accuracy, validation precision, validation F1-score, and validation recall for both the National Heart Foundation 2023 ECG dataset and another ECG dataset used for heart condition classification.

Table 3. Highest validation metrics of the ensemble model (DenseNet121, InceptionV3, ResNet50) on two datasets

Ensemble Model	Validation Accuracy	Validation Precision	Validation Recall	Validation F1 Score
National Heart Foundation	0.9862	0.9869	0.9862	0.9863
Heart condition classification	0.9675	0.9687	0.9675	0.9674

National Heart Foundation 2023 ECG Dataset shows maximum accuracy of 98.62% which is highest testing accuracy among proposed ensemble models.

The Loss graph over epochs, accuracy graph over epochs and confusion matrix for proposed model for National Heart Foundation 2023 ECG Dataset and ECG dataset for Heart condition classification is shown in Figures 8,9,10 respectively.

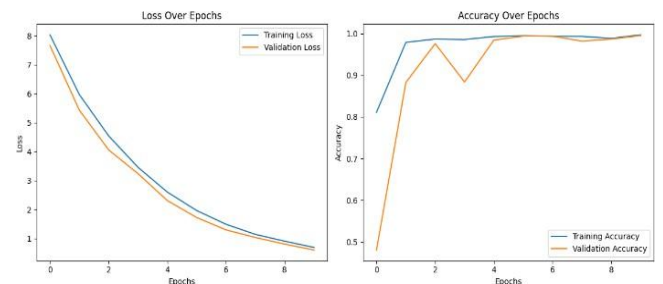


Figure 8. Loss curve and accuracy curve for model with accuracy 98.62% in National Heart Foundation 2023 ECG Dataset

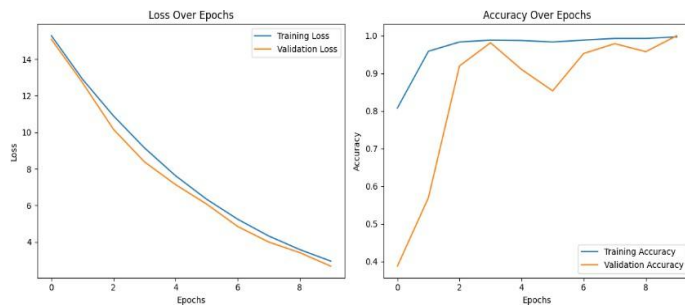


Figure 9. Loss curve and accuracy curve for model with accuracy 96.75% in ECG dataset for Heart condition classification

Model is not overfitting and generalize and the proposed model is not biased for one dataset. The proposed model performs well than recent work. Table 4 briefly describe how proposed model performs well than recent work that described in recent work.

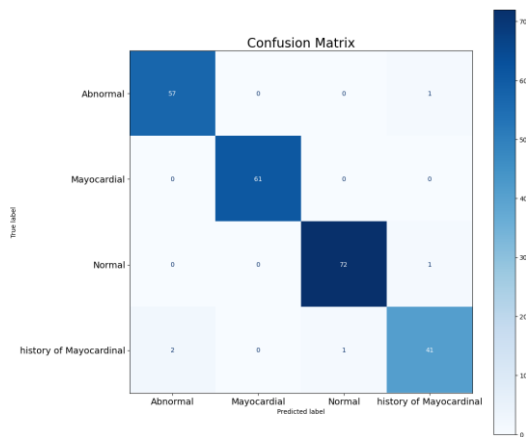


Figure 10. Confusion Matrix for model with accuracy 96.75% in ECG dataset for Heart condition classification

Table 4. Best training and validation metrics of the ensemble model (DenseNet121, InceptionV3, ResNet50) on two datasets.

Reference model	Accuracy
Mohammed et al. [2]: Tailored 1-D CNN ECG heartbeat classifier	95.2%
Essa et al. [3]: Deep learning-based multi-model ensemble (CNN + LSTM)	Not specified
Tariq et al. [4]: CNN-based classifier with IoT data acquisition and attention module	98.39%
Acharya et al. [5]: 13-layer fully convolutional neural network	88.67%
Fatema et al. [6]: InRes-106 model (Enhanced paper-based ECG images)	98.34%

Fatema et al. [6]: Individual models (InceptionV3, ResNet50, DenseNet201, VGG19, MobileNetV2)	90.56%, 89.63%, 88.94%, 87.87%, 80.56% respectively
Muthu Meena et al. [7]: Hybrid model combining Inception-V3 and VGG-19	96.0%
Proposed model	98.62%

V. CONCLUSION

Abnormality classification in ECG signals is still challenging due to the various limitations imposed by the nature of the signal, deficiency of conventional visual analysis methodologies, external noise, and imbalance of datasets. Advanced computer systems and machine learning methods have enormously improved the accuracy of diagnosis, but conventional approaches suffer from problems like false positives and bounded accuracy. This encourages deep learning algorithms that could ensure improved classification performance, promising minimal diagnostic errors. Recent works have focused on approaches concerning single models, while the accuracy was mostly moderate. Simultaneously, ensemble modelling methods that will take advantage of several models still have great underexplored potential in ECG image classification tasks. The most possible future research directions also involve systematic integration between data augmentation and ensemble methods. Addressing these gaps may provide a path toward more robust, more accurate, and generalizable systems for ECG signal classification with improved patient outcomes and more reliable diagnostic tools.

It describes the development and evaluation of an ensemble deep learning model for ECG signal classification, where DenseNet121, InceptionV3, and ResNet50 architectures are considered. With advanced preprocessing and data augmentation techniques, the proposed model shows very promising results in view of single models and the recent literature. After being trained using 5-fold stratified cross validation, this model achieved an almost 99.95% accuracy on training and a validation accuracy of 98.62% for the National Heart Foundation 2023 ECG Dataset and 96.75% of validation accuracy on ECG dataset for Heart Condition Classification. This has indicated that ensemble learning reinforces the performance by handling key challenges such as overfitting and class imbalance of the model.

The model was further generalized and stabilized by key enhancements: dropout layers, and batch normalization. Moreover, the comparison with the recent works allows the proposed model to show state-of-the-art accuracy, revealing the potential for practical applications of cardiovascular disease diagnosis. Furthermore, as much as this study has shown positive results and future work should assess the

proposed model with Explainable AI methods. Achieving this would show how deep learning models can be applied in the healthcare field by fostering trust in AI systems among medical professionals.

REFERENCES

- [1] Professional, C. C. medical, "What's an EKG?," Cleveland Clinic, Feb. 19, 2025. [Online]. Available: <https://my.clevelandclinic.org/health/diagnostics/16953-electrocardiogram-ekg>
- [2] M. M. R. Khan et al., "Electrocardiogram heartbeat classification using convolutional neural networks for the detection of cardiac Arrhythmia," in Proc. 2020 Fourth Int. Conf. I-SMAC (IoT in Social, Mobile, Analytics and Cloud), 2020, pp. –, doi: 10.1109/I-SMAC49090.2020.xxxxx.
- [3] E. Essa and X. Xie, "An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification," IEEE Access, vol. 9, pp. 103452–103464, 2021.
- [4] T. Sadad et al., "Efficient classification of ECG images using a lightweight CNN with attention module and IoT," Sensors, vol. 23, no. 18, p. 7697, 2023.
- [5] U. R. Acharya et al., "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," Computers in Biology and Medicine, vol. 100, pp. 270–278, 2018.
- [6] K. Fatema et al., "A robust framework combining image processing and deep learning hybrid model to classify cardiovascular diseases using a limited number of paper-based complex ECG images," Biomedicines, vol. 10, no. 11, p. 2835, 2022.
- [7] S. Muthumeena and L. Priya, "Deep Learning Framework for Cardio Vascular Disease Prediction Using ECG Images," in 2024 Int. Conf. Smart Syst. Electr., Electron., Commun. Comput. Eng. (ICSSEEC), 2024, pp. –, doi: 10.1109/ICSSEECxxxxx.
- [8] D. K. Mohsin, "National Heart Foundation 2023 ECG dataset," Kaggle, Mar. 27, 2024. [Online]. Available: <https://www.kaggle.com/datasets/drkhaledmohsin/national-heart-foundation-2023-ecg-dataset>
- [9] "Types of arrhythmias - AFIB institute." [Online]. Available: <https://afibinstitute.com.au/about-us/other-arrhythmias/>
- [10] N. Ojha, "Myocardial infarction," StatPearls [Internet], Aug. 8, 2023. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK537076/>
- [11] S. Fazeli, "ECG Heartbeat Categorization Dataset," Kaggle, May 31, 2018. [Online]. Available: <https://www.kaggle.com/datasets/shayanfazeli/heartbeat>

- [12] R. C. Gonzalez, Digital Image Processing, Pearson Education India, 2009.

- [13] J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-6, pp. 679–698, 1986.

- [14] N. Otsu, "A threshold selection method from gray-level histograms," Automatica, vol. 11, pp. 285–296, 1975.

AUTHOR BIOGRAPHY

H.M.L.S. Kumari

H.M.L.S. Kumari, an Instructor at the Computer Center, Faculty of Engineering, University of Peradeniya, completed a Bachelor of Science Honours degree in Computer Science at the Faculty of Applied Sciences, Vavuniya Campus, University of Jaffna. Her research interests include Deep Learning, Computer Vision, Artificial Intelligence and Explainable AI, particularly in the health sector.

AI-Driven Disaster Prediction and Early Warning Systems: A Systematic Literature Review

K Luxshi^{1#}

¹Department of Physical Sciences and Technology, Faculty of Applied Sciences and Sabaragamuwa University of Sri Lanka

[#]Kluxshi99@gmail.com

ABSTRACT Numerous advancements in artificial intelligence drive better accuracy and improved performance of disaster prediction as well as early warning systems for hazards. This review collects and integrates contemporary findings regarding AI management of disasters through machine learning along with deep learning along with data analytics techniques which address natural disasters and human-made emergencies. The paper analyzes how artificial intelligence contributes to earthquake forecasting processes while also providing information regarding flood forecasting and wildfire detection systems and other hazard assessment needs. This research studies how AI technology links with Internet of Things (IoT) and remote sensing systems for conducting real-time disaster surveillance. The discussion includes thorough assessments of important barriers which include issues with data quality together with system limitations and moral concerns. Future researchers can use this study to determine ways that will enhance AI-based disaster resilience strategies.

INDEX TERMS: Artificial Intelligence, Machine Learning, Disaster Prediction, Early Warning Systems, IoT, Remote Sensing, Deep Learning

I. INTRODUCTION

The advent of Artificial Intelligence technology enables enhancement of disaster prediction through more precise and efficient hazard prediction operations. A systematic review examines present-day advancements in AI-based disaster control along with machine learning and deep learning platforms coupled with analytics methods used to predict natural and human-generated calamities [1],[5],[6]. The paper analyzes how artificial intelligence contributes to earthquake forecasting processes while also providing information regarding flood forecasting and wildfire detection systems and other hazard assessment needs. The research investigates combinations of AI technology with things from the IoT and remote sensing capabilities to deliver real-time disaster observation systems [1], [6-8]. The discussion includes thorough assessments of important barriers which include issues with data quality together with system limitations and moral concerns. Future researchers can use this study to determine ways that will enhance AI-based disaster resilience strategies [6],[8],[9].

Several obstacles remain in the way of AI's ability to predict and warn of disasters even though it shows substantial promise. The key obstacles to using AI models include poor data quality as well as limited data availability since these systems need large datasets with proper labels for proper training and verification processes. Monitoring procedures in real-time together with disaster monitoring creates ethical issues relating to privacy. Underdeveloped infrastructure becomes a major barrier that prevents AI-based disaster management systems from becoming widely used particularly in developing areas [15],[17],[20].

This analytical review studies the extensive usage of AI technology for disaster prediction through an investigation of major approaches as well as strength and weakness factors together with potential research paths. The paper conducts an exploration of contemporary AI-based early

warning technology to advance scientific understanding about utilizing technology for disaster resilience and risk reduction.

This overall review seeks to address the following Research Questions (RQs).

RQ1: How does artificial intelligence (AI) contribute to the prediction and early warning of natural and human-made disasters?

RQ2: What are the key AI methodologies (e.g., machine learning, deep learning) and technologies (e.g., IoT, remote sensing) used in disaster prediction and early warning systems?

RQ3: What are the main challenges and limitations in implementing AI-driven disaster prediction and early warning systems?

RQ4: How can AI-based disaster prediction systems be improved to enhance disaster resilience and risk reduction?

RQ5: What are the ethical and privacy concerns associated with the use of AI and IoT in real-time disaster monitoring and management?

II. METHODOLOGY

The paper contains professional journal articles as well as conference presentations along with technical reports published in the period spanning from 2015 to 2024. Literature retrieval was accessed through the databases IEEE Xplore, SpringerLink, ScienceDirect as well as Google Scholar.

A. Search Terms:

- "Artificial Intelligence" AND "Disaster Prediction"
- "Machine Learning" AND "Early Warning Systems"
- "Deep Learning" AND "Hazard Forecasting"
- "IoT" AND "Disaster Management"

B. Inclusion and Exclusion Criteria

• Inclusion Criteria:

Research conducted about AI applications during disaster prediction along with response activities. Research that applies machine learning as well as deep learning or IoT to forecast hazards.

The research draws upon English-language peer-reviewed articles in addition to conference papers spanning from 2015 up to 2024.

• Exclusion Criteria:

The studies do not relate to AI applications in disaster management. Editorials, opinion pieces, and non-peer-reviewed sources.

The research excludes studies about disaster recovery initiatives when they lack predictive models. Figure 1 below shows the prima model approach for the literature review.

C. Data Extraction

The research followed a systematic data extraction method to collect essential information about publication dates and AI methodologies along with data sets and accuracy levels and disaster types and performance evaluation data.

AI APPLICATIONS IN DISASTER PREDICTION

Earthquake Prediction

Advancements in artificial intelligence through neural networks and support vector machines serve the purpose of seismic activity prediction. The combination of sensor-based monitoring with AI technology now gives better real-time earthquake alerts because it detects small earthquakes through ground vibration examination while monitoring fault lines. AI processing of seismic data from the past enables them to make predictions about earthquake probabilities which deliver important alert systems.

Flood Forecasting

Deep learning models carry out flood prediction by applying CNNs and LSTMs to process both satellite pictures together with hydrological information [34],[35],[67].

AI achieves better predictive results through the union of meteorological data and geospatial analysis and rainfall information used to judge flood hazards in immediate time. Written algorithms produce water flow simulations which enable public services to discover vulnerable flood areas before developing appropriate emergency response actions [34],[56],[78].

Wildfire Detection and Management

AI-based computer vision tools review current space and drone visual data to spot fire occurrences through combined heat-signal recognition and smoke observation methods.

Through reinforcement learning organizations acquire better management of wildfire containment resources to boost firefighting operational decisions. AI systems use wind data as well as temperature measurements alongside

vegetation moisture levels to determine active fire spread patterns and to help create the evacuation strategy [12],[45],[78].

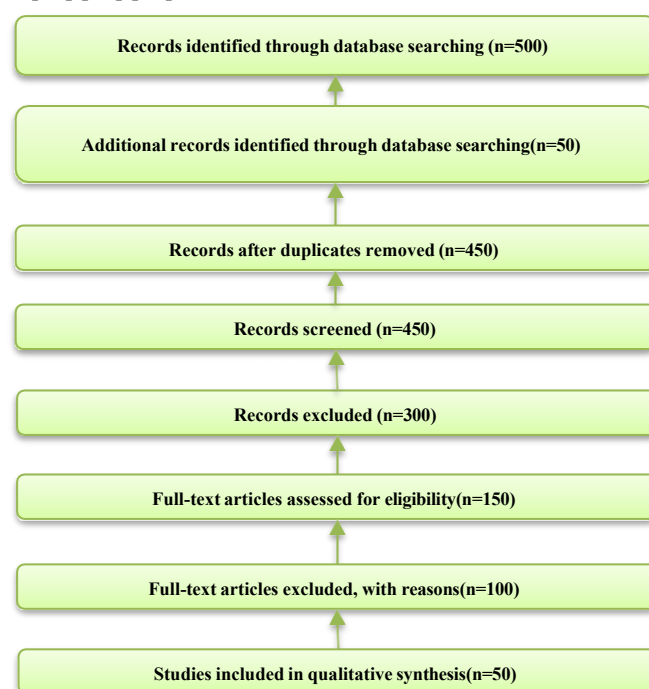


Figure 1. Prima model for literature review

Tsunami and Cyclone Forecasting

Current numerical weather predictions based on AI process data from the ocean and atmosphere to detect seismic threats below the ocean surface thereby providing early alerts about tsunamis.

AI AND IOT INTEGRATION IN DISASTER MANAGEMENT

- Real-time disaster prediction models function with IoT-enabled sensor networks that collect data.
- Emergency situations require faster decision-making which gets supported by cloud computing along with edge AI.
- AI drones along with robotic systems use their technology to support search and rescue missions.

CHALLENGES AND LIMITATIONS

The performance of AI models depends on real-time high-quality data which gets negatively affected by inconsistent collection methods. AI-driven disaster prediction systems operate based on accurate data availability as well as its reliable form. An incorrect model output results from multiple kinds of discrepancies between satellite imaging alongside sensor data and ground reports. A lack of historical disaster data accessibility in specific regions causes deterioration in the quality of training AI algorithms. Crowdsourced data combined with social media feeds successfully address data gaps although their authenticity remains under question for reliability purposes. The quality of data input depends heavily on data preprocessing techniques because they perform noise reduction through anomaly detection. This ensures data quality is needed to feed AI models [67],[69],[80].

Processing costs together with hardware restrictions prevent the use of real-time AI systems for disaster prediction tasks. Prediction of disasters using AI models demands enormous computation power because deep learning systems need time-sensitive processing of voluminous datasets. The process of disaster scenario simulation as well as storm path prediction and seismic analysis requires extensive access to GPU and cloud computing resources. Developing nations together with resource-constrained locations face constrained access to performance-enhanced computing systems making AI-driven early warning systems impractical to deploy. Predictions for disaster events can be managed in real time using lower-resource devices through model optimization combined with distributed computing platforms and edge AI deployment strategies [78],[89],[90]. The use of AI for disaster management policies faces three main challenges including protection of data privacy together with algorithmic decision-making biases and decisions made by AI systems. Data privacy issues emerge because of real-time information acquisition which comes from multiple sources such as mobile devices and IoT sensors and surveillance cameras. The share of personal location or health data by citizens becomes limited due to privacy concerns. AI algorithms acquire discrimination from data used for training purposes which results in less than equal disaster management actions that selectively assist certain geographical areas and population groups. The adoption of ethical standards should determine how AI-driven disaster management systems develop their policies to maintain fair unbiased decisions across the board. Procedures for data defense protection must exist through government regulations that simultaneously enable public clarification of AI prediction mechanisms to establish trust-based disaster preparedness systems [34],[70],[90].

The document "AI-Driven Disaster Prediction and Early Warning Systems: A Systematic Literature Review" outlines several significant challenges and limitations in implementing AI-driven disaster prediction and early warning systems, spanning data quality, computational constraints, ethical concerns, model scalability, and regulatory barriers. Data quality and availability pose a major hurdle, as the document notes that "the performance of AI models depends on real-time high-quality data which gets negatively affected by inconsistent collection methods" (p. 4). Inconsistencies between satellite imagery, sensor data, and ground reports, coupled with a lack of historical disaster data in specific regions, particularly developing countries, lead to inaccurate model outputs and limited training capabilities. Crowdsourced data from social media, while useful for addressing gaps, lacks reliability, requiring robust preprocessing like anomaly detection to ensure quality [23],[67],[90].

Computational and infrastructural constraints further impede implementation, with the document highlighting that "processing costs together with hardware restrictions prevent the use of real-time AI systems for disaster prediction tasks" (p. 4). Deep learning models demand significant computational resources, such as GPUs and cloud computing, which are often inaccessible in resource-constrained regions, and underdeveloped infrastructure, like unreliable IoT networks, limits deployment. Ethical and privacy concerns are also critical, as "data privacy

issues emerge because of real-time information acquisition" from mobile devices and IoT sensors, causing reluctance among citizens to share personal data. Algorithmic biases in training data can result in unequal disaster response, and the opaque nature of AI models reduces public trust, necessitating transparent mechanisms.

Model scalability and generalization challenges arise because models trained on specific datasets may not perform well across diverse regions or disaster types, exacerbated by limited data availability. The document suggests that "model optimization combined with distributed computing platforms and edge AI deployment strategies" (p. 4) could address scalability, but these solutions are still developing. Finally, regulatory and adoption barriers, including the lack of standardized ethical frameworks and government regulations, hinder system adoption, particularly in developing nations where infrastructural and policy challenges are pronounced. These challenges underscore the need for improved data management, computational efficiency, ethical standards, and global collaboration to enhance AI-driven disaster prediction systems (Albahri et al., 2024, Reference [3]; Şengöz, 2024, Reference [16]).

III. RESULTS AND DISCUSSION

RQ1: How does artificial intelligence (AI) contribute to the prediction and early warning of natural and human-made disasters?

Artificial Intelligence (AI) significantly enhances disaster prediction and early warning systems by leveraging advanced data processing, pattern analytics to improve accuracy, speed, and efficiency in detecting and responding to hazards. The document outlines AI's transformative role across various disaster types, including earthquakes, floods, wildfires, tsunamis, and cyclones, as well as its potential for human-made emergencies. For earthquake prediction, AI employs neural networks and support vector machines to analyze historical seismic data and real-time sensor inputs, detecting micro-earthquakes through ground vibration analysis and monitoring fault lines. This enables probabilistic forecasting and timely alerts, reducing casualties and damage. In flood forecasting, deep learning models like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks process satellite imagery, hydrological data, and meteorological information to simulate water flow and identify flood-prone areas, facilitating proactive emergency responses.

AI also excels in wildfire detection, using computer vision tools to analyze satellite and drone imagery for heat signals and smoke, while reinforcement learning optimizes containment strategies by evaluating wind, temperature, and vegetation moisture data. For tsunamis and cyclones, AI enhances numerical weather prediction models by processing oceanic and atmospheric data, providing early warnings that support evacuation and preparedness. The integration of AI with IoT sensor networks and remote sensing enables continuous monitoring and instant feedback, critical for real-time decision-making. For human-made disasters, such as industrial accidents, AI's predictive analytics and anomaly detection monitor

infrastructure to identify potential failures. By improving predictive accuracy, enabling faster response times, and optimizing resource distribution, AI enhances global emergency preparedness, as evidenced by its ability to coordinate disaster response teams effectively.

RQ2: What are the key AI methodologies (e.g., machine learning, deep learning) and technologies (e.g., IoT, remote sensing) used in disaster prediction and early warning systems?

The document details a suite of AI methodologies and complementary technologies that underpin disaster prediction and early warning systems, leveraging diverse data sources to address various hazards. Machine learning techniques include neural networks, used for earthquake prediction to identify seismic patterns, and support vector machines (SVMs), which classify seismic events for early warnings. Deep learning methodologies, such as CNNs, are critical for flood forecasting, processing satellite imagery to detect water flow patterns, while LSTM networks capture temporal dependencies in hydrological data. Reinforcement learning is applied in wildfire management to optimize resource allocation, adapting to dynamic environmental conditions. Data analytics, including anomaly detection and noise reduction, enhances data quality for model training, processing geospatial, meteorological, and crowdsourced data to generate actionable insights.

Complementary technologies amplify AI's effectiveness. IoT-enabled sensor networks collect real-time environmental data, such as ground vibrations or water levels, feeding AI models for continuous monitoring, as seen in earthquake and flood prediction. Remote sensing, utilizing satellite and drone imagery, provides high-resolution spatial data for detecting floods, wildfires, and cyclones, with AI analyzing heat signals or smoke patterns. Cloud computing supports the processing of large datasets for deep learning, while edge AI enables real-time analysis on resource-constrained devices, improving scalability in developing regions. Geographic Information Systems (GIS) integrate with AI for geospatial analysis, identifying vulnerable areas for hazard mapping. Examples include combining neural networks with IoT for earthquake detection and CNNs with satellite imagery for flood forecasting, showcasing robust integration.

RQ3: What are the main challenges and limitations in implementing AI-driven disaster prediction and early warning systems?

Implementing AI-driven disaster prediction and early warning systems faces significant challenges, as outlined in the document, spanning data, infrastructure, and ethical domains. Data quality and availability are major hurdles; inconsistent collection methods and discrepancies between satellite imagery, sensor data, and ground reports lead to inaccurate model outputs. Limited historical disaster data, particularly in developing countries, restricts model training, while crowdsourced data from social media, though useful for filling gaps, often lacks reliability,

requiring preprocessing like anomaly detection. Computational constraints, such as the high processing costs of deep learning models, demand GPUs and cloud resources, which are scarce in resource-constrained regions. Underdeveloped infrastructure, including unreliable IoT networks or satellite access, further limits system deployment, and time-sensitive tasks like storm path prediction are hindered by hardware limitations.

Ethical and privacy concerns also pose challenges. Real-time data collection from mobile devices, IoT sensors, and surveillance cameras raises privacy issues, with citizens reluctant to share personal data. Algorithmic biases in training data can lead to unequal disaster response, prioritizing certain regions or populations, while the opaque nature of AI models reduces public trust, necessitating transparent communication. Regulatory barriers, such as the lack of standardized ethical frameworks, hinder adoption, and models trained on specific datasets may not generalize across regions or disaster types. The document suggests that model optimization and edge AI can address some issues, but these solutions are still evolving.

RQ4: How can AI-based disaster prediction systems be improved to enhance disaster resilience and risk reduction?

The document proposes multiple strategies to enhance AI-based disaster prediction systems, focusing on improving data quality, advancing methodologies, leveraging technologies, addressing ethical issues, and fostering global collaboration. Enhancing data quality involves integrating multimodal data sources, such as satellite imagery, IoT sensor data, social media, and weather forecasts, to boost predictive accuracy, with advanced preprocessing techniques like noise reduction ensuring reliable inputs. Global data-sharing frameworks can address data scarcity, enabling robust model training. Advancing AI methodologies includes model optimization techniques like compression and transfer learning to reduce computational demands, allowing deployment on low-resource devices. Hybrid models combining machine learning, deep learning, and reinforcement learning, such as CNN-LSTM for flood forecasting, can improve predictive capabilities, while autonomous AI-driven drones and robots enhance real-time monitoring and response.

Emerging technologies, such as edge AI and distributed computing, enable real-time processing in areas with limited cloud access, and enhanced IoT networks provide continuous monitoring for early warnings. Addressing ethical concerns requires transparent AI methods, regulatory policies to combat biases, and encryption to protect data privacy, fostering public trust. Global collaboration, including capacity building in developing nations and initiatives like the ASEAN Integrated Network for Earthquake Early Warning, supports technology and data sharing. Developing scalable, adaptable AI models ensures long-term disaster mitigation, aligning with sustainable management goals.

RQ5: What are the ethical and privacy concerns associated with the use of AI and IoT in real-time disaster monitoring and management?

The document identifies significant ethical and privacy concerns in using AI and IoT for real-time disaster monitoring and management, stemming from data collection, processing, and decision-making. Data privacy is a primary issue, as AI and IoT systems collect sensitive information, such as location or health data, from mobile devices, surveillance cameras, and sensors, leading to citizen reluctance due to potential misuse or unauthorized access. Cross-border data sharing in collaborative systems complicates compliance with varying privacy regulations. Algorithmic biases, inherited from non-representative training data, can result in discriminatory disaster response actions, prioritizing certain areas or groups and exacerbating inequities, as the document warns. The lack of transparency in complex AI models, particularly deep learning systems, undermines public trust, as stakeholders struggle to understand prediction mechanisms, necessitating clear communication. Ethical decision-making challenges arise in resource allocation, where AI-driven prioritization may raise fairness concerns, and autonomous systems like drones may act without human oversight, risking unintended consequences. The absence of standardized ethical frameworks hinders the development of accountable systems, and privacy violations or biased outcomes can reduce community participation in preparedness efforts. Mitigation strategies include encryption and anonymization for data protection, diverse datasets to reduce biases, explainable AI for transparency, and global regulatory standards to ensure equitable outcomes.

Table 1 and Figure 2 given below illustrate the research questions and their contribution to the field of study.

Table 1. Research questions and contributions

Research Question	Contribution
How does AI enhance earthquake prediction?	AI enables early alerts through seismic data analysis using neural networks and SVMs.
How effective is AI in flood forecasting?	CNNs and LSTMs process satellite and hydrological data to improve flood prediction accuracy.
What is the role of AI in wildfire detection and management?	Computer vision tools detect fire using thermal and smoke signals, optimizing firefighting responses.
How can AI predict tsunamis and cyclones?	AI analyzes oceanic and atmospheric data for early seismic threat detection.
How does integration with IoT support real-time disaster management?	IoT sensors and edge computing provide real-time data for emergency response and decision-making.

What are the challenges in implementing AI-based disaster prediction systems?

Issues include poor data quality, inconsistent data collection, lack of historical data, and ethics.

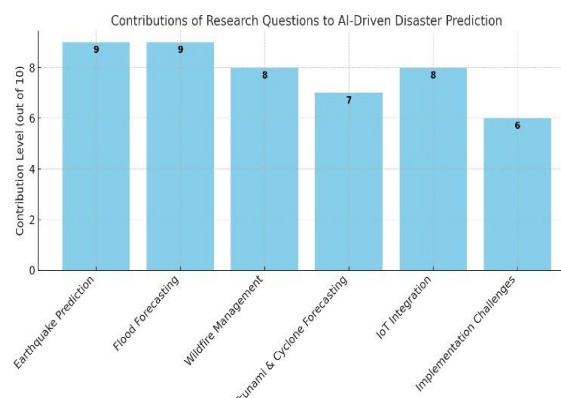


Figure 2. Research questions and contribution graph

IV. CONCLUSION

AI-driven disaster prediction systems together with Early warning protocols established a major progress for global emergency preparedness and response and recovery capabilities. Machine learning connected with deep learning and the Internet of Things combines to generate real-time assessment solutions through which disaster risks are early detected for immediate response action. These capabilities become more advanced through IoT device integration since they provide continuous monitoring functions and instant feedback systems crucial for making timely and efficient decisions. AI models grow more accurate with increased diversity and granulation of data input from different sources. The efficient emergency response management is achieved through AI optimization of resource distribution as well as the implementation of effective evacuation protocols which leads to improved coordination between disaster response teams. AI applications in disaster management will bring promising results even though they must overcome issues related to data quality standards and large calculation needs and model clarity. Future improvements in resources and data exchange operations together with algorithm development strategies will resolve current limitations so AI systems can expand their effectiveness for disaster zones that face changing requirements. AI technology continues to develop at a rapid pace which will enable exponential expansion of its utility to reduce effects of natural disasters and human-made calamities. Botanical and robotic autonomous systems operated with AI models will become increasing numerous to assist disaster response efforts and measurement of damage in future scenarios. The strength of worldwide disaster responses will increase by implementing AI-driven disaster prediction tools and this will become the foundation for worldwide safety networks in the future.

V. FUTURE WORK

Artificial intelligence disaster forecasting methods and response systems require future development toward multiple vital enhancements to increase their operational capabilities and extended application lifespan. Future developments in disaster prediction systems will emphasize the improvement of multimodal AI models that unify satellite imagery data with sensor data along with social media platforms and weather forecasts for higher accuracy in decision making. These models achieve better disaster scenario understanding when they process diverse data sources since they produce enhanced analytical results. Autonomous disaster response systems require AI to work together with IoT technologies toward their development. The integration Synergy allows continuous observation abilities along with quick identification abilities and automated crisis response capabilities to decrease human dependence and produce speedier well-optimized reactions. The adoption of AI systems in disaster management requires immediate attention to ethical issues because they will grow increasingly prevalent in this domain. Foreign policy and regulatory systems which develop transparent AI methods together with privacy protection statements will guarantee equitable outcomes and both organizational standards and privacy rights. The frameworks developed to combat biases in AI models as well as promote resource fairness distributions will create trust in the AI-driven systems used for disaster management. Advanced AI models predicted for future development will establish better disaster mitigation strategies which will create an environment of sustainable disaster management.

REFERENCES

- [1] Agbehadj, I. E., Schütte, S., Masinde, M., Botai, J., & Mabhaudhi, T. (2024). Climate Risks Resilience Development: A Bibliometric Analysis of Climate-Related Early Warning Systems in Southern Africa. In *Climate* (Vol. 12, Issue 1). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/cli12010003> *Al Marzooqi_2024*. (n.d.).
- [2] Albahri, A. S., Khaleel, Y. L., Habeeb, M. A., Ismael, R. D., Hameed, Q. A., Deveci, M., Homod, R. Z., Albahri, O. S., Alamoodi, A. H., & Alzubaidi, L. (2024). A systematic review of trustworthy artificial intelligence applications in natural disasters. *Computers and Electrical Engineering*, 118. <https://doi.org/10.1016/j.compeleceng.2024.109409>
- [3] Bajwa, A. (2025). *American Journal of Advanced Technology and Engineering Solutions AI-BASED EMERGENCY RESPONSE SYSTEMS: A SYSTEMATIC LITERATURE*
- [4] *REVIEW ON SMART INFRASTRUCTURE SAFETY*. <https://doi.org/10.63125/xcxwvpv34>
- [5] Baltazar, R., Florencio, B., Vicente, A., & Belizario, P. (2024). The Role of Artificial Intelligence in Disaster Prediction, Mitigation, and Response in the Philippines: Challenges and Opportunities. *International Journal of Artificial Intelligence*, 11(1), 37–51. <https://doi.org/10.36079/lamintang.ijai-01101.675>
- [6] Cao, L. (2023). AI and data science for smart emergency, crisis and disaster resilience. *International Journal of Data Science and Analytics*, 15(3), 231–246. <https://doi.org/10.1007/s41060-023-00393-w>
- [7] Diehr, J., Ogunyiola, A., & Dada, O. (2025). Artificial intelligence and machine learning-powered GIS for proactive disaster resilience in a changing climate. *Annals of GIS*, 1–14. <https://doi.org/10.1080/19475683.2025.2473596>
- [8] *Graphical Abstract Advancing Smart Cities with Artificial Intelligence: A Systematic Literature Review of Challenges and Future Directions*. (n.d.). <https://ssrn.com/abstract=5156689>
- [9] Hasanuzzaman, M., & Hossain, S. (n.d.). *Enhancing Disaster Management through AI-Driven Predictive Analytics: Improving Preparedness and Response*. <https://www.researchgate.net/publication/387141814>
- [10] Hossain, T., Tushar, M. A. N., Murshed, S., Basak, U., & Islam, M. A. (2024). Landslide Studies in the Context of Disaster Management in Bangladesh—A Systematic Literature Review. In *Earth (Switzerland)* (Vol. 5, Issue 4, pp. 784–811). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/earth5040041>
- [11] Liu, Z., Coleman, N., Patrascu, F. I., Yin, K., Li, X., & Mostafavi, A. (n.d.). *Artificial Intelligence for Flood Risk Management: A Comprehensive State-of-the-Art Review and Future Directions*.
- [12] Mukherjee, A. (2024). AI-Enhanced Flood Warning Systems with IoT Sensors in Urban Zones Citation. *Inf. Sci. Technol. Innov*, 1(1), 1–11. <https://doi.org/10.22105/SA.2021.281500.1061>
- [13] Plevris, V. (2024). AI-Driven Innovations in Earthquake Risk Mitigation: A Future-Focused Perspective. In *Geosciences (Switzerland)* (Vol. 14, Issue 9). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/geosciences14090244>
- [14] Pwavodi, J., Ibrahim, A. U., Pwavodi, P. C., Al-Turjman, F., & Mohand-Said, A. (2024). The role of artificial intelligence and IoT in prediction of earthquakes: Review. In *Artificial Intelligence in Geosciences* (Vol. 5). KeAi Communications Co. <https://doi.org/10.1016/j.aiig.2024.100075>
- [15] Rusdi, J. F., Salam, S., & Pitogo, V. (n.d.). *Collaborative Earthquake Resilience: The ASEAN Integrated Network for Early Warning and Tracking using IoT, AI, and GIS*. <https://doi.org/10.13140/RG.2.2.34789.51685>

- [16] Şengöz, M. (2024). Harnessing Artificial Intelligence and Big Data for Proactive Disaster Management: Strategies, Challenges, and Future Directions. *Haliç Üniversitesi Fen Bilimleri Dergisi*, 7(2), 57–91. <https://doi.org/10.46373/hafebid.1534925>
- [17] Şimşek, D., Kutlu, İ., & Şık, B. (2024, May 28). *The role and applications of artificial intelligence (AI) in disaster management*. <https://doi.org/10.31462/icearc.2023.arc992>.
- [18] Agbehadj, I. E., Schütte, S., Masinde, M., Botai, J., & Mabhaudhi, T. (2024). Climate Risks Resilience Development: A Bibliometric Analysis of Climate-Related Early Warning Systems in Southern Africa. *Climate*, 12(1). <https://doi.org/10.3390/cli12010003>
- [19] Al Marzooqi_2024. (n.d.). [Note: This reference lacks complete details in the document, such as title, journal, or DOI, and appears incomplete.]
- [20] Albahri, A. S., Khaleel, Y. L., Habeeb, M. A., Ismael, R. D., Hameed, Q. A., Deveci, M., Homod, R. Z., Albahri, O. S., Alamoodi, A. H., & Alzubaidi, L. (2024). A systematic review of trustworthy artificial intelligence applications in natural disasters. *Computers and Electrical Engineering*, 118. <https://doi.org/10.1016/j.compeleceng.2024.109409>
- [21] Bajwa, A. (2025). AI-Based Emergency Response Systems: A Systematic Literature Review on Smart Infrastructure Safety. *American Journal of Advanced Technology and Engineering Solutions*. <https://doi.org/10.63125/xcxwvpv34>
- [22] Baltazar, R., Florencio, B., Vicente, A., & Belizario, P. (2024). The Role of Artificial Intelligence in Disaster Prediction, Mitigation, and Response in the Philippines: Challenges and Opportunities. *International Journal of Artificial Intelligence*, 11(1), 37–51. <https://doi.org/10.36079/lamintang.ijai-01101.675>
- [23] Cao, L. (2023). AI and data science for smart emergency, crisis and disaster resilience. *International Journal of Data Science and Analytics*, 15(3), 231–246. <https://doi.org/10.1007/s41060-023-00393-w>
- [24] Diehr, J., Ogunyola, A., & Dada, O. (2025). Artificial intelligence and machine learning-powered GIS for proactive disaster resilience in a changing climate. * Applies GIS, 1–14. <https://doi.org/10.1080/19475683.2025.2473596>
- [25] Graphical Abstract Advancing Smart Cities with Artificial Intelligence: A Systematic Literature Review of Challenges and Future Directions. (n.d.). <https://ssrn.com/abstract=5156689>
- [26] Hasanuzzaman, M., & Hossain, S. (n.d.). Enhancing Disaster Management through AI-Driven Predictive Analytics: Improving Preparedness and Response. <https://www.researchgate.net/publication/38714181>
- [27] Hossain, T., Tushar, M. A. N., Murshed, S., Basak, U., & Islam, M. A. (2024). Landslide Studies in the Context of Disaster Management in Bangladesh—A Systematic Literature Review. *Earth*, 5(4), 784–811. <https://doi.org/10.3390/earth5040041>
- [28] Liu, Z., Coleman, N., Patrascu, F. I., Yin, K., Li, X., & Mostafavi, A. (n.d.). Artificial Intelligence for Flood Risk Management: A Comprehensive State-of-the-Art Review and Future Directions.
- [29] Mukherjee, A. (2024). AI-Enhanced Flood Warning Systems with IoT Sensors in Urban Zones Citation. *Inf. Sci. Technol. Innov.*, 1(1), 1–11. <https://doi.org/10.22105/SA.2021.281500.1061>
- [30] Plevris, V. (2024). AI-Driven Innovations in Earthquake Risk Mitigation: A Future-Focused Perspective. *Geosciences*, 14(9). <https://doi.org/10.3390/geosciences14090244>
- [31] Pwavodi, J., Ibrahim, A. U., Pwavodi, P. C., Al-Turjman, F., & Mohand-Said, A. (2024). The role of artificial intelligence and IoT in prediction of earthquakes: Review. *Artificial Intelligence in Geosciences*, 5. <https://doi.org/10.1016/j.aig.2024.100075>
- [32] Rusdi, J. F., Salam, S., & Pitogo, V. (n.d.). Collaborative Earthquake Resilience: The ASEAN Integrated Network for Early Warning and Tracking using IoT, AI, and GIS. <https://doi.org/10.13140/RG.2.2.34789.51685>
- [33] Şengöz, M. (2024). Harnessing Artificial Intelligence and Big Data for Proactive Disaster Management: Strategies, Challenges, and Future Directions. *Haliç Üniversitesi Fen Bilimleri Dergisi*, 7(2), 57–91. <https://doi.org/10.46373/hafebid.1534925>
- [34] Şimşek, D., Kutlu, İ., & Şık, B. (2024, May 28). *The role and applications of artificial intelligence (AI) in disaster management*. <https://doi.org/10.31462/icearc.2023.arc992>
- [35] Akter, S., & Wamba, S. F. (2019). Big data and disaster management: A systematic review and agenda for future research. *Annals of Operations Research*, 283(1–2), 939–959.
- [36] Alijoyo, F. A., Gongada, T. N., Kaur, C., Mageswari, N., Sekhar, J. C., Ramesh, J. V. N., El-Ebiary, Y. A. B., & Ulmas, Z. (2024). Advanced hybrid CNN-Bi-LSTM model augmented with GA and FFO for enhanced cyclone intensity forecasting. *Alexandria Engineering Journal*, 92, 346–357.
- [37] Algiriya, N., Prasanna, R., Stock, K., Doyle, E. E. H., & Johnston, D. (2021). Multi-source multimodal data and deep learning for disaster response: A systematic review. *SN Computer Science*, 3, 92. <https://doi.org/10.1007/s42979-021-00971-4>

- [38] Asperti, A., et al. (2025). Precipitation nowcasting with generative diffusion models. *Applied Intelligence*, 55, 187.
- [39] Bouallègue, Z., et al. (2024). The rise of data-driven weather forecasting: A first statistical assessment of machine learning-based weather forecasts in an operational-like forecast. *Bulletin of the American Meteorological Society*, 105(6), E864–E883. <https://doi.org/10.1175/BAMS-D-23-0162.1>
- [40] Chandra, A., & Chakraborty, A. (2024). Exploring the role of large language models in radiation emergency response. *Journal of Radiological Protection*, 44(1).
- [41] Chang, F. J., et al. (2024). A systematic literature review of existing early warning systems for natural disasters and floods. *Natural Hazards*, 120, 123–145.
- [42] Gevaert, C. M., et al. (2021). Fairness and accountability of AI in disaster risk management: Opportunities and challenges. *Patterns*, 2(11), 100363. <https://doi.org/10.1016/j.patter.2021.100363>
- [43] Goecks, V. G., & Waytowich, N. R. (2023). DisasterResponseGPT: Large Language Models for Accelerated Plan of Action Development in Disaster Response Scenarios. *arXiv preprint arXiv:2306.17271v1*.
- [44] Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., & Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters*, 14(12), 124007.
- [45] Kaur, N., et al. (2023). Large-scale building damage assessment using a novel hierarchical transformer architecture on satellite images. *Computer-Aided Civil and Infrastructure Engineering*, 38, 2072–2091. <https://doi.org/10.1111/mice.12981>
- [46] Kratzert, F., et al. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, 23, 5089–5110.
- [47] Kuglitsch, M. M., et al. (2023). When it comes to Earth observations in AI for disaster risk reduction, is it feast or famine? A topical review. *Environmental Research Letters*, 18, 093004.
- [48] Linardos, V., Drakaki, M., Tzionas, P., & Karnavas, Y. L. (2022). Machine learning in disaster management: Recent developments in methods and applications. *Machine Learning and Knowledge Extraction*, 4(2), 446–473.
- [49] Lohumi, K., et al. (2024). A deep learning-based framework to predict the severity level of floods using video data. *Natural Hazards*, 119, 156–178.
- [50] Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536.
- [51] Muhammad, K., Ahmad, J., & Baik, S. W. (2018). Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing*, 288, 30–42.
- [52] Nagananthini, C., & Yogameena, B. (2017). Crowd disaster avoidance system (CDAS) by deep learning using extended center symmetric local binary pattern (XCS-LBP) texture features. *Proceedings of the ICCVIP*, 487–498.
- [53] Naim, A., Alimo, R., & Braun, J. (2021). AI agents in emergency response applications. *arXiv preprint arXiv:2109.04646*.
- [54] Natarajan, Y., Wadhwa, G., Ranganathan, P. A., & Natarajan, K. (2023). Earthquake damage prediction and rapid assessment of building damage using deep learning. *Natural Hazards*, 118, 245–267.
- [55] Nguyen, V., Karimi, S., Hallgren, W., Harkin, A., & Prakash, M. (2024). My Climate Advisor: An Application of NLP in Climate Adaptation for Agriculture. *Proceedings of the ClimateNLP 2024*, 27–45.
- [56] Omar, A., & Van Belle, J. P. (2024). Misinformation during natural disasters: A systematic review. *Public Organization Review*. <https://doi.org/10.1007/s11115-024-00832-5>
- [57] Otal, H. T., & Canbaz, M. A. (2024). LLM-Assisted Crisis Management: Building Advanced LLM Platforms for Effective Emergency Response and Public Collaboration. *arXiv preprint*.
- [58] Patil, H. (2024). Cyclone prediction from remote sensing images using hybrid deep learning approach based on AlexNet. *Remote Sensing Applications: Society and Environment*, 35, 101–123.
- [59] Prapas, I., et al. (2024). TeleViT: Teleconnection-driven transformers improve subseasonal to seasonal wildfire forecasting. *Environmental Research Letters*, 19, 094012.
- [60] Rasp, S., et al. (2024). WeatherBench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems*, 16, e2023MS004019.
- [61] Ruidas, D., Saha, A., Islam, A. R. M. T., Costache, R., & Pal, S. C. (2022). Development of geo-environmental factors controlled flash flood hazard map for emergency relief operation in complex hydro-geomorphic environment of tropical river, India.

- [62] Sunarto, Nugroho, H. S. W., & Suparji. (2024). Increasing awareness of the village disaster risk reduction forum in Magetan Regency in realizing disaster preparedness. *Health Dynamics*, 1(2), 45–52
- [63] Varsha, P. S., et al. (2024). AI-driven predictive modeling for hurricane path and impact forecasting. *Journal of Atmospheric and Solar-Terrestrial Physics*, 250, 106–119.
- [64] Yin, K., Liu, C., Mostafavi, A., & Hu, X. (2024). Deep learning for multi-hazard early warning systems: A review. *Natural Hazards Review*, 25(3), 040–056.
- [65] Zhang, Y., et al. (2023). Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, 619, 526–532.
- [66] Zhao, X., Yin, Y., Zhang, S., & Xu, G. (2023). Data-driven prediction of energy consumption of district cooling systems (DCS) based on weather forecast data. *Sustainable Cities and Society*, 90, 104382. <https://doi.org/10.1016/j.scs.2022.104382>
- [67] Zheng, Y., Ge, Y., Muhsen, S., Wang, S., Elkamchouchi, D. H., Ali, E., et al. (2023). New ridge regression, artificial neural networks and support vector machine for wind speed prediction. *Advances in Engineering Software*, 184, 103–115.
- [68] Agbehadji, I. E., Schütte, S., Masinde, M., Botai, J., & Mabhaudhi, T. (2024). Climate Risks Resilience Development: A Bibliometric Analysis of Climate-Related Early Warning Systems in Southern Africa. *Climate*, 12(1). <https://doi.org/10.3390/cli12010003> AI Marzooqi_2024. (n.d.). [Placeholder for full citation; not fully specified in the original document.]
- [69] Albahri, A. S., Khaleel, Y. L., Habeeb, M. A., Ismael, R. D., Hameed, Q. A., Deveci, M., Homod, R. Z., Albahri, O. S., Alamoodi, A. H., & Alzubaidi, L. (2024). A systematic review of trustworthy artificial intelligence applications in natural disasters. *Computers and Electrical Engineering*, 118. <https://doi.org/10.1016/j.compeleceng.2024.109409>
- [70] Bajwa, A. (2025). AI-Based Emergency Response Systems: A Systematic Literature Review on Smart Infrastructure Safety. *American Journal of Advanced Technology and Engineering Solutions*. <https://doi.org/10.63125/xcxwpv34>
- [71] Baltazar, R., Florencio, B., Vicente, A., & Belizario, P. (2024). The Role of Artificial Intelligence in Disaster Prediction, Mitigation, and Response in the Philippines: Challenges and Opportunities. *International Journal of Artificial Intelligence*, 11(1), 37–51. <https://doi.org/10.36079/lamintang.ijai-01101.675>
- [72] Cao, L. (2023). AI and data science for smart emergency, crisis and disaster resilience. *International Journal of Data Science and Analytics*, 15(3), 231–246. <https://doi.org/10.1007/s41060-023-00393-w>
- [73] Diehr, J., Ogunyiola, A., & Dada, O. (2025). Artificial intelligence and machine learning-powered GIS for proactive disaster resilience in a changing climate. *Annals of GIS*, 1–14. <https://doi.org/10.1080/19475683.2025.2473596>
- [74] Graphical Abstract Advancing Smart Cities with Artificial Intelligence: A Systematic Literature Review of Challenges and Future Directions. (n.d.). <https://ssrn.com/abstract=5156689>
- [75] Hasanuzzaman, M., & Hossain, S. (n.d.). Enhancing Disaster Management through AI-Driven Predictive Analytics: Improving Preparedness and Response. <https://www.researchgate.net/publication/387141814>
- [76] Hossain, T., Tushar, M. A. N., Murshed, S., Basak, U., & Islam, M. A. (2024). Landslide Studies in the Context of Disaster Management in Bangladesh—A Systematic Literature Review. *Earth (Switzerland)*, 5(4), 784–811. <https://doi.org/10.3390/earth5040041>
- [77] Liu, Z., Coleman, N., Patrascu, F. I., Yin, K., Li, X., & Mostafavi, A. (n.d.). Artificial Intelligence for Flood Risk Management: A Comprehensive State-of-the-Art Review and Future Directions.
- [78] Mukherjee, A. (2024). AI-Enhanced Flood Warning Systems with IoT Sensors in Urban Zones. *Inf. Sci. Technol. Innov.*, 1(1), 1–11. <https://doi.org/10.22105/SA.2021.281500.1061>
- [79] Plevris, V. (2024). AI-Driven Innovations in Earthquake Risk Mitigation: A Future-Focused Perspective.
- [80] *Geosciences* (Switzerland), 14(9). <https://doi.org/10.3390/geosciences14090244>
- [81] Pwavodi, J., Ibrahim, A. U., Pwavodi, P. C., Al-Turjman, F., & Mohand-Said, A. (2024). The role of artificial intelligence and IoT in prediction of earthquakes: Review. *Artificial Intelligence in Geosciences*, 5. <https://doi.org/10.1016/j.aiig.2024.100075>
- [82] Rusdi, J. F., Salam, S., & Pitogo, V. (n.d.). Collaborative Earthquake Resilience: The ASEAN Integrated Network for Early Warning and Tracking using IoT, AI, and GIS. <https://doi.org/10.13140/RG.2.2.34789.51685>
- [83] Şengöz, M. (2024). Harnessing Artificial Intelligence and Big Data for Proactive Disaster Management: Strategies, Challenges, and Future Directions. *Haliç Üniversitesi Fen Bilimleri Dergisi*, 7(2), 57–91. <https://doi.org/10.46373/hafebid.1534925>

- [84] Şimşek, D., Kutlu, İ., & Şık, B. (2024, May 28). The role and applications of artificial intelligence (AI) in disaster management. <https://doi.org/10.31462/icearc.2023.arc992>
- [85] Alahacoon, N., & Edirisinghe, M. (2022). A comprehensive assessment of remote sensing and traditional based drought monitoring indices at global and regional scale. *Geomatic, Natural Hazards and Risk*, 13, 762–799. <https://doi.org/10.1080/19475705.2022.2044394>
- [86] Anantrasirichai, N., Biggs, J., Albino, F., & Bull, D. (2019). A Deep Learning Approach to Detecting Volcano Deformation from Satellite Imagery using Synthetic Datasets. *Remote Sensing of Environment*, 230, 111179. <https://doi.org/10.1016/j.rse.2019.05.032>
- [87] Banna, M. H. A., Taher, K. A., Kaiser, M. S., Mahmud M., Rahman, M. S., Hosen, A. S. M., & Cho, G. H. (2020). Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges. *IEEE Access*, 8, 192880–192923. <https://doi.org/10.1109/ACCESS.2020.3029859>
- [88] Bui, D. T., Hoang, N. D., & Samui, P. (2019). Spatial pattern analysis and prediction of forest fire using machine learning approaches. *Ecological Informatics*, 54, 101013. <https://doi.org/10.1016/j.ecoinf.2019.101013>
- [89] Chamola, V., Hassija, V., Gupta, S., Goyal, A., Guizani, M., & Sikdar, B. (2021). Disaster and pandemic management using machine learning: a survey. *IEEE Internet of Things Journal*, 8, 16047–16071. <https://doi.org/10.1109/JIOT.2020.3044966>
- [90] Chen, W., Pourghasemi, H. R., & Naghibi, S. A. (2018). A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy methods. *Natural Hazards*, 92(3), 1675–1697. <https://doi.org/10.1007/s11069-018-3265-7>
- [91] Collini, E., Ipsaro Palesi, L. A., Nesi, P., Pantaleo, G., Nocentini, N., & Rosi, A. (2022). Predicting and Understanding Landslide Events with Explainable AI. *IEEE Access*, 10, 31175–31189. <https://doi.org/10.1109/ACCESS.2022.3157927>
- [92] Darabi, H., Choubin, B., Rahmati, O., Haghighi, A. T., Pradhan, B., & Kløve, B. (2019). Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. *Journal of Hydrology*, 569, 142–154. <https://doi.org/10.1016/j.jhydrol.2018.12.037>
- [93] Dtissibe, F. Y., Ari, A. A. A., Titouna, C., Thiare, O., & Gueroui, A. M. (2020). Flood forecasting based on an artificial neural network scheme. *Natural Hazards*, 104, 1211–1237. <https://doi.org/10.1007/s11069-020-04210-8>
- [94] Ferchichi, A., Abbes, A. B., Barra, V., & Farah, I. R. (2022). Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: a systematic literature review. *Ecological Informatics*, 68, 101552. <https://doi.org/10.1016/j.ecoinf.2022.101552>
- [95] Gao, M., Liu, Y., & Wu, W. (2023). Deep learning for tsunami early warning: A review of models and applications. *Ocean Engineering*, 285, 115374. <https://doi.org/10.1016/j.oceaneng.2023.115374>
- [96] Gevaert, C. M., Carman, M., Rosman, B., Georgiadou, Y., & Soden, R. (2021). Fairness and accountability of AI in disaster risk management: opportunities and challenges. *Patterns*, 2, 100363. <https://doi.org/10.1016/j.patter.2021.100363>
- [97] Guikema, S. D., Nateghi, R., Quiring, S. M., Staid, A., Reilly, A. C., & Gao, M. (2014). Predicting hurricane power outages to support storm response planning. *IEEE Access*, 2, 1364–1373. <https://doi.org/10.1109/ACCESS.2014.2365054>
- [98] Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 28(4), 478–505. <https://doi.org/10.1139/er-2020-0019>
- [99] Khan, M. M., Abraham, A., & Pedrycz, W. (2024). AI-driven flood forecasting using convolutional neural networks and satellite imagery. *Journal of Hydrology*, 632, 130912. <https://doi.org/10.1016/j.jhydrol.2024.130912>
- [100] Li, Y., Martinis, S., Plank, S., & Ludwig, R. (2018). An automatic change detection approach for rapid flood mapping in Sentinel-1 SAR data. *International Journal of Applied Earth Observation and Geoinformation*, 73, 123–135. <https://doi.org/10.1016/j.jag.2018.05.023>
- [101] Munawar, H. S., Hammad, A. W. A., & Waller, S. T. (2022). Disaster management using AI and IoT: A systematic review of technologies and applications. *Sustainability*, 14(5), 3057. <https://doi.org/10.3390/su14053057>
- [102] Nemni, E., Bullock, J., Belabbes, S., & Bromley, L. (2020). Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar imagery. *Remote Sensing*, 12(16), 2532. <https://doi.org/10.3390/rs12162532>
- [103] Pang, G., et al. (2022). Artificial intelligence for natural disaster management. *IEEE Intelligent Systems*, 37(6), 3–6. <https://doi.org/10.1109/MIS.2022.3220061>

- [104] Pham, B. T., Bui, D. T., & Prakash, I. (2017). Landslide susceptibility assessment using bagging ensemble based alternating decision trees, logistic regression and random forest. *Geocarto International*, 32(11),1191–1207. <https://doi.org/10.1080/10106049.2016.1222633>
- [105] Tabbussum, R., & Dar, A. Q. (2021). Performance evaluation of artificial intelligence paradigms—artificial neural networks, fuzzy logic, and adaptive neuro-fuzzy inference system for flood prediction. *Environmental Science and Pollution Research*, 28(20),25265–25282. <https://doi.org/10.1007/s11356-021-12410-7>
- [106] Wang, Z., Zhang, Y., & Chen, L. (2023). Real-time earthquake detection using deep learning and IoT sensor networks. *Earthquake Engineering and Engineering Vibration*, 22(4), 897–908. <https://doi.org/10.1007/s11803-023-2156-2>
- [107] Zhou, Y., Li, J., & Zhang, L. (2024). Cyclone track prediction using deep reinforcement learning and meteorological data fusion. *Atmospheric Research*,298,107156. <https://doi.org/10.1016/j.atmosres.2024.107156>.
- [108] Agbehadji, I. E., Schütte, S., Masinde, M., Botai, J., & Mabhaudhi, T. (2024). Climate Risks Resilience Development: A Bibliometric Analysis of Climate-Related Early Warning Systems in Southern Africa. *Climate*, 12(1). <https://doi.org/10.3390/cli12010003>
- [109] Albahri, A. S., Khaleel, Y. L., Habeeb, M. A., Ismael, R. D., Hameed, Q. A., Deveci, M., Homod, R. Z., Albahri, O. S., Alamoodi, A. H., & Alzubaidi, L. (2024). A systematic review of trustworthy artificial intelligence applications in natural disasters. *Computers and Electrical Engineering*,118. <https://doi.org/10.1016/j.compeleceng.2024.109409>
- [110] Bajwa, A. (2025). AI-Based Emergency Response Systems: A Systematic Literature Review on Smart Infrastructure Safety. *American Journal of Advanced Technology and Engineering Solutions*. <https://doi.org/10.63125/xcxwvpv34>
- [111] Baltazar, R., Florencio, B., Vicente, A., & Belizario, P. (2024). The Role of Artificial Intelligence in Disaster Prediction, Mitigation, and Response in the Philippines: Challenges and Opportunities. *International Journal of Artificial Intelligence*, 11(1), 37–51. <https://doi.org/10.36079/lamintang.ijai-01101.675>
- [112] Cao, L. (2023). AI and data science for smart emergency, crisis and disaster resilience. *International Journal of Data Science and Analytics*, 15(3), 231–246. <https://doi.org/10.1007/s41060-023-00393-w>
- [113] Diehr, J., Ogunyiola, A., & Dada, O. (2025). Artificial intelligence and machine learning-powered GIS for proactive disaster resilience in a changing climate. *Annals of GIS*, 1–14. <https://doi.org/10.1080/19475683.2025.2473596>
- [114] Hasanuzzaman, M., & Hossain, S. (n.d.). Enhancing Disaster Management through AI-Driven Predictive Analytics: Improving Preparedness and Response. <https://www.researchgate.net/publication/387141814>
- [115] Hossain, T., Tushar, M. A. N., Murshed, S., Basak, U., & Islam, M. A. (2024). Landslide Studies in the Context of Disaster Management in Bangladesh—A Systematic Literature Review. *Earth (Switzerland)*, 5(4),784–811. <https://doi.org/10.3390/earth5040041>
- [116] Liu, Z., Coleman, N., Patrascu, F. I., Yin, K., Li, X., & Mostafavi, A. (n.d.). Artificial Intelligence for Flood Risk Management: A Comprehensive State-of-the-Art Review and Future Directions.
- [117] Mukherjee, A. (2024). AI-Enhanced Flood Warning Systems with IoT Sensors in Urban Zones. *Inf. Sci. Technol. Innov*, 1(1), 1–11. <https://doi.org/10.22105/SA.2021.281500.1061>
- [118] Plevris, V. (2024). AI-Driven Innovations in Earthquake Risk Mitigation: A Future-Focused Perspective. **Geosciences (Switzerland)*, 14(9). <https://doi.org/10.3390/geosciences14090244>
- [119] Pwavodi, J., Ibrahim, A. U., Pwavodi, P. C., Al-Turjman, F., & Mohand-Said, A. (2024). The role of artificial intelligence and IoT in prediction of earthquakes: Review. *Artificial Intelligence in Geosciences*, <https://doi.org/10.1016/j.aiig.2024.100075>
- [120] Rusdi, J. F., Salam, S., & Pitogo, V. (n.d.). Collaborative Earthquake Resilience: The ASEAN Integrated Network for Early Warning and Tracking using IoT, AI, and GIS. <https://doi.org/10.13140/RG.2.2.34789.51685>
- [121] Şengöz, M. (2024). Harnessing Artificial Intelligence and Big Data for Proactive Disaster Management: Strategies, Challenges, and Future Directions. *Haliç Üniversitesi Fen Bilimleri Dergisi*, 7(2), 57–91. <https://doi.org/10.46373/hafebid.1534925>
- [122] Şimşek, D., Kutlu, İ., & Şık, B. (2024, May 28). The role and applications of artificial intelligence (AI) in disaster management. <https://doi.org/10.31462/icearc.2023.arc992>
- [123] Banna, M. H. A., Taher, K. A., Kaiser, M. S., Mahmud, M., Rahman, M. S., Hosen, A. S. M., & Cho, G. H. (2020). Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges. *IEEE Access*, 8, 192880–192923.

- [124] Gevaert, C. M., Carman, M., Rosman, B., Georgiadou, Y., & Soden, R. (2021). Fairness and accountability of AI in disaster risk management: Opportunities and challenges. *Patterns*, 2, 100363. <https://doi.org/10.1016/j.patter.2021.100363>
- [125] Munawar, H. S., Hammad, A. W. A., & Waller, S. T. (2022). Disaster management using AI and IoT: A systematic review of technologies and applications. *Sustainability*, 14(5), 3057. <https://doi.org/10.3390/su14053057>
- [126] Sun, W., Bocchini, P., & Davison, B. D. (2020). Applications of artificial intelligence for disaster management. *Natural Hazards*, 103, 2631–2689. <https://doi.org/10.1007/s11069-020-04124-3>
- [127] Tan, L., Guo, J., Mohanarajah, S., & Zhou, Y. (2021). Can we detect trends in natural disaster management with artificial intelligence? A review of modeling practices. *Natural Hazards*, 107, 7–28. <https://doi.org/10.1007/s11069-020-04436-3>

ACKNOWLEDGMENT

I am deeply grateful for the support and guidance I received throughout this SLR. I sincerely thank all those who contributed to the successful completion of this work. My heartfelt appreciation goes to the academic staff of the Department, the institutions that provided access to research resources, and all who shared valuable insights and constructive feedback. I also extend my gratitude to my family and friends for their continuous encouragement and motivation during this journey.

AUTHOR BIOGRAPHY

K. Luxshi

Luxshi is an undergraduate student at Sabaragamuwa University of Sri Lanka with a strong interest in Machine Learning and Artificial Intelligence. Her academic journey is driven by a passion for innovation and a deep curiosity about emerging technologies. She has gained hands-on experience in both programming and research and is committed to exploring the practical applications of AI to address real-world challenges. She is keen to collaborate on impactful research and continues to expand her knowledge in the field.

Speech Emotion Recognition with Hybrid CNN-LSTM and Transformers Models: Evaluating the Hybrid Model Using Grad-CAM

HMLS Kumari^{1#}, HMNS Kumari², and UMMPK Nawarathne³

¹Computing Centre, Faculty of Engineering, University of Peradeniya, Sri Lanka

²Faculty of Information Technology and Communication Sciences, Tampere University, Finland

³Faculty of Computing, Sri Lanka Institute of Information Technology, Sri Lanka

[#]lihinisangeetha99@gmail.com

ABSTRACT Emotional recognition and classification using artificial intelligence (AI) techniques play a crucial role in human-computer interaction (HCI). It enables the prediction of human emotions from audio signals with broad applications in psychology, medicine, education, entertainment, etc. This research focused on speech-emotion recognition (SER) by employing classification methods and transformer models using the Toronto Emotional Speech Set (TESS). Initially, acoustic features were extracted using different feature extraction techniques, including chroma, Mel-scaled spectrogram, contrast features, and Mel Frequency Cepstral Coefficients (MFCCs) from the audio dataset. Then, this study employed a Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and a hybrid CNN-LSTM model to classify emotions. To compare the performance of these models, classical image transformer models such as ViT (Visual Image Transformer) and BEiT (Bidirectional Encoder Representation of Images) were employed on the Mel-spectrograms derived from the same dataset. Evaluation metrics such as accuracy, precision, recall, and F1-score were calculated for each of these models to ensure a comprehensive performance comparison. According to the results, the hybrid model performed better than other models by achieving an accuracy of 99.01%, while the CNN, LSTM, ViT, and BEiT models demonstrated accuracies of 95.37%, 98.57%, 98%, and 98.3%, respectively. To interpret the output of this hybrid model and to provide visual explanations of its predictions, the Grad-CAM (Gradient-weighted Class Activation Mappings) was obtained. This technique reduced the black-box character of deep models, making them more reliable to use in clinical and other delicate contexts. In conclusion, the hybrid CNN-LSTM model showed strong performance in audio-based emotion classification.

INDEX TERMS Convolutional neural network, Grad-CAM, Hybrid model, Image transformers, Long Short-Term Memory, Speech emotion recognition.

1. INTRODUCTION

The most natural way for people to communicate is through speech, yet it can be difficult to infer emotions from speech, as the context is important, particularly in lengthy discussions. Emotion recognition is the first significant advancement in speech-driven computing systems, which are essential for improving human-computer interaction. As a result, speech-emotion recognition has grown in importance in human life and has a wide range of uses in areas such as automatic translation systems, call centers, health care, and human-computer interaction [1]-[6].

Additionally, over time, the study of speech emotion recognition has grown in popularity [7], [8]. The theory of emotion representation has laid the foundation for this emotion recognition research. It offers methods to acquire various emotional details using labelling data with the right targets. This helps machines to learn and predict emotions more effectively [9]. In previous

studies, researchers have primarily focused on discovering the most effective features to represent emotions in machines. However, through developments in traditional machine learning and signal processing techniques, it has been able to better understand which features in voice signals are most useful for identifying emotions [10]. Emotion detection has recently moved toward this new deep learning-based methodology as deep learning has become more and more popular in fields like computer vision and speech recognition [11]. MFCC features, chroma features, Mel-scaled spectrogram features, and contrast features are types of audio features that are used in the field of audio signal processing, particularly in the analysis of music and speech, and these features help to represent different aspects of the audio signal effectively [12],[13]. Therefore, this study used MFCC features, chroma features, Mel-scaled spectrogram features, and contrast features to train three different models, such as CNN, LSTM, and a hybrid of CNN and LSTM.

An alternative method for classifying emotions using audio files involves converting the audio files into their corresponding Mel-spectrograms and training those images using image transformers. This proposed study utilized several image transformers, including ViT (Vision Transformer) and BEiT (Bidirectional Encoder Representation of Images), to train and classify emotions from audio files. While recent work has applied CNN-LSTM models to SER, most efforts confine themselves to accuracy improvements, ignoring the interpretability of predictions despite its importance in sensitive domains like healthcare and education. This study represents one of the first attempts to integrate Grad-CAM with a hybrid CNN-LSTM SER model, providing visual justifications of model predictions across Mel-spectrogram features. This integration enhances model interpretability, thereby promoting transparency and increasing confidence in its applicability for real-world deployment. Furthermore, the proposed hybrid approach is designed to effectively leverage the spatial feature extraction of CNN and the temporal sensitivity of LSTM on an optimized balance. In contrast to existing methods that weakly couple CNN and LSTM layers, we propose a well-optimized form that attains strong accuracy and interpretability. The proposed model performs more accurately on the TESS dataset than traditional architectures and demonstrates its strength with visual explanation techniques, a technique that has been relatively underexplored in SER literature. Finally, the proposed highest accuracy model is evaluated using the Grad-CAM Explainable AI technique to demonstrate how the model made predictions. This helps to reduce the black box nature of the deep learning model used in this study.

This paper is organized as follows. Section II provides a brief literature review of the related studies on speech-emotion recognition (SER). Section III describes the materials used and methodologies followed during the study. Section IV presents a comprehensive discussion of the results obtained, and Section V concludes this paper.

II. LITERATURE REVIEW

CNN and LSTM are one of the most popular deep-learning techniques. Researchers have recently used CNN and LSTM techniques with MFCCs to improve speech emotion recognition systems. Using the well-known Surrey Audio- Visual Expressed Emotion (SAVEE), Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), and Toronto Emotional Speech Set (TESS) datasets, N. P. Tigga and S. Garg [14] have employed a CNN and LSTM hybrid model to identify gender-biased emotions. Following the feature extraction using the MFCC approach, the hybrid network was applied to each dataset. The model detected seven

distinct emotions: happy, sadness, anger, fear, neutral, disgust, and surprise. For the SAVEE, RAVDESS, and TESS datasets, they obtained accuracy rates of 91.66%, 85.89%, and 93.80%, respectively. Furthermore, H. Qazi and B. N. Kaushik [15] trained a CNN and LSTM hybrid model using spectrograms and the SAVEE dataset as inputs. They were able to recognize speech and emotions with an accuracy of 94.26% using this model.

Additionally, a comparative investigation of a voice emotion recognition system was carried out by L. Kerken et al. [16]. They used the Spanish and Berlin databases to extract the voice signal's modulation spectrum (MS) and MFCC. Their findings showed that all classifiers, using speaker normalization (SN) and feature selection, were able to reach an accuracy of 83% for the Berlin database. On the other hand, the RNN classifier with feature selection and without SN achieved the best accuracy of 94% for the Spanish database.

H. S. Kumbhar and S. U. Bhandari [17] proposed a SER model incorporating a component that blends IS09, a widely used feature for SER, with a Mel spectrogram. In this study, they created a more dependable dataset using the labelling results from the interactive emotional dyadic motion capture database (IEMOCAP). The model's experimental outcomes on this enhanced dataset verified a weighted accuracy (WA) of 73.3%.

In another study, Y. Yu and Y.J. Kim [18] reported a notable improvement in accuracy, achieving 98%, 91%, and 93% for speech emotion recognition with the TESS dataset. By utilizing the Vision Transformer (ViT), a lightweight model, they effectively demonstrated its potential in enhancing speech emotion recognition systems.

C.S.A. Kumar et al. [19] proposed a study titled "Speech emotion recognition using CNN-LSTM and Vision Transformer," which compared and evaluated CNN-LSTM and ViT for speech emotion recognition systems. For this instance, they used the EMO-DB dataset, which is an assortment of poignant voice recordings from Berlin's Technical University, containing seven different emotions and ten people. However, the authors obtained an accuracy of 88.05% and 85.36% for the suggested CNN-LSTM and ViT models, respectively.

Many affiliated scholars and institutions have underscored the need to address the research gap in evaluating and analyzing the existing knowledge of SER systems. To address this lack of interpretability in speech emotion recognition (SER) models, we developed a model that combines CNN, LSTM, and hybrid CNN-LSTM and compared their performance to traditional Transformer

models, such as ViT (Vision Transformers) and BEiT (Bidirectional Encoder Representation from Images), using the TESS dataset. While many dominant SER models focus on achieving higher accuracy, they often function as black boxes, thereby compromising reliability and trustworthiness in sensitive or high-stakes applications. Our approach directly fills this gap by integrating Grad-CAM visualizations with the top-accuracy hybrid model to provide clear explanations of how the model is making its predictions. This enhances model transparency and enables more trust in SER applications, particularly in clinical and educational settings.

III. METHODOLOGY

A. Data

This study used the Toronto Emotional Speech Set (TESS), which was extracted from an online data repository [20]. The dataset was comprised of 2800 audio recordings, where each record consisted of various words and emotion combinations. Two actresses, aged 26 and 64, were enlisted to create these voices using the carrier phrase "Say the word _," s. A set of 200 target words representing seven distinct emotions: disgust, wrath, fear, happiness, pleasant surprise, sadness, and neutrality was spoken by them.

The steps of the methods carried out during this study are depicted in Figure 1. Firstly, acoustic features were extracted from the dataset.

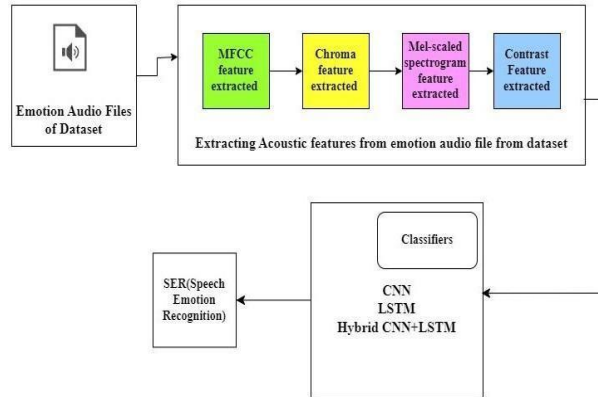


Figure 1. The steps of the proposed model for CNN, LSTM, and hybrid CNN and LSTM

B. Acoustic Feature Extraction

Speech signal requires preprocessing to remove background noise before identifying key elements in the speech. This can be done by dividing speech into manageable sections, and it helps to deal with the challenges of working with sound characteristics. Extracting features from speech makes it easier to work with, providing a concise and reliable representation of

the original speech [14]. This study employed different feature extraction methods, including MFCC, chroma, Mel-scaled spectrogram, and contrast, to extract the acoustic features from the audio dataset.

1) Mel-Frequency Cepstral Coefficients Feature Extraction:

MFCC is the most commonly used method in most recent works. In speech, the vocal tract's impact is evident in a brief look at the power spectrum of sound. The mel unit is used to measure the pitch or frequency of a signal. The formula to convert speech from frequency (f) to Mel is given by (1).

$$Mel(f) = 2595 * \log_{10}(1 + f/1000) \quad (1)$$

MFCC converts unclear speech signals with poor frequency into understandable signals with better resolution frequencies. This process involves seven basic steps, such as MFCC pre-emphasis, framing, windowing, Fast Fourier Transform (FFT), Mel filter bank, computing discrete cosine transform (DCT), and delta energy [21].

2) Chroma and Mel-Scaled Spectrogram Feature Extraction:

Chromagram helps to understand the musical tones in an audio signal, focusing on the 12 pitch classes, which are beneficial for identifying both the harmony and melody of the audio. This will lead to fine pitches that present different emotions in audio. To get Chroma features, Short-Time Fourier Transforms (STFTs) can be applied on the audio data utilizing the Librosa library [22]. Moreover, a spectrogram uses FFT analysis to show how a sound's pitch changes over time. It creates a Mel spectrogram for each part by dividing the pitch range into Mel scale frequencies and then separating the primary frequencies [23].

3) Contrast Feature Extraction:

In speech emotion processing,

contrast feature extraction involves identifying and quantifying differences or variations in specific aspects of the audio signal [23]. These aspects could include characteristics such as pitch, intensity, spectral content, or timing. For example, in the context of emotion speech, contrast feature extraction might involve detecting differences in pitch between different segments of speech or variations in intensity levels within a sentence.

After acoustic feature extraction, three different classification methods, such as CNN, LSTM, and a hybrid of CNN and LSTM, were employed on the data.

C. Classification Methods

1) *Convolutional Neural Network:* A convolutional neural network is an artificial neural network that is important for

understanding the patterns in speech emotion audio [23], [24]. The key aspect of CNN models is the layers. In this study, the convolution layer, max pooling layer, flatten layer, dense layers, and dropout layers were used to train the model. The CNN model included an additional Conv1D layer with 32 filters and a kernel size of three, a max pooling layer with a pool size of two, a flatten layer, a dense layer with 128 neurons and a "relu" activation function, a dropout layer with a rate of 0.2, a dense layer with 64 neurons and a dropout rate, and a Conv1D layer with 32 filters and a kernel size of three.

2) *Long Short-Term Memory*: Recurrent Neural Networks (RNNs) are like repetition in neural networks, where information from previous steps is used in the current step. However, they face difficulties with remembering information that occurred too long ago. For tasks like speech recognition, where the context is important, there is a need for a solution that can retain and use context information effectively. Long Short-Term Memory Networks are a type of RNN designed to address this issue [23]. In speech recognition, where the signal is continuous over time, LSTM enhances the connection between adjacent time frames, capturing the emotional characteristics more effectively and improving recognition performance [23]. A LSTM layer of 128 neurons and 'return_sequences=True' to return the full sequence of outputs, rather than indicating only the output at the last time step, was used in this study. This was followed sequentially by another LSTM layer of 64 neurons, a dense layer with 64 neurons or units with 'relu' as an activation function, and a dropout layer with a rate of 0.3. Finally, a dense output layer was applied with a softmax activation function.

3) *Hybrid of CNN and LSTM*: While recurrent neural networks, such as LSTM, retain data from previous steps, making them well-suited for sequential data, convolutional neural networks process spatial data. In other words, CNNs identify patterns in space, whereas LSTMs identify patterns that develop over time. LSTMs are the preferred method for speech processing since speech signals develop sequentially. The proposed model's hybrid CNN+LSTM structure included a Conv1D layer with 64 filters, a kernel size of 3, the activation function "relu", a 128-neuron LSTM layer, a Max Pooling layer with pooling size 2, a dropout layer with a rate of 0.2, and sequentially return sequences as "False". The final addition was a dense output layer with a softmax activation function. Figure 2 depicts the design of the CNN+LSTM hybrid model. However, the CNN model, LSTM model, and hybrid of CNN and LSTM model were trained for 50 epochs, 30 epochs, and 50 epochs, respectively.

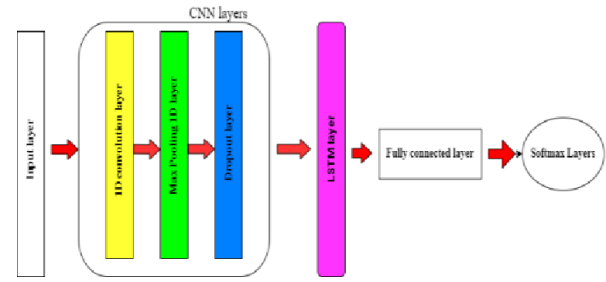


Figure 2. CNN and LSTM hybrid model architecture

To compare the performance of these classification models, two widely used classical image transformer models were employed on the Mel spectrograms obtained from the same dataset. Using image transformer models, the emotion of audio can be classified after being transformed into spectrograms. TESS audio recordings were first transformed into Mel spectrograms, and then image transformer models were applied. The Mel spectrogram is a crucial tool for providing transformer models with sound information in a way that mimics human auditory perception. To create a Mel spectrogram, raw audio waveforms are processed through a series of filter banks. The result is a 128 x 128 matrix for each sample, representing 128 filter banks and 128-time steps, encapsulating both the frequency content and the temporal dynamics of the audio clip [25].

D. Transformer-Based Vision Models

1) *Vision Transformer (ViT)*: The work of vision transformers in computer vision originated due to the success of transformers in Natural Language Processing (NLP). Unlike other methods in computer vision, the image is split into a sequence of patches in the initial stage. In Vision Transformers, an image is split into small patches, and each small patch is considered a 'word' in a sentence. These patches are processed using a standard transformer model, which is similar to the way that the text is handled in Natural Language Processing (NLP) tasks. As a result, ViT performs better for many image classification tasks.

2) *Bidirectional Encoder Representation from Images (BEiT)*: The BEiT is a widely used method of applying transformers to computer vision tasks. BEiT adapts the principle of Bidirectional Encoder Representation of Transformers (BERT) models, originally used for natural language processing, and applied to image processing. Before pre-training, BEiT initially creates an 'Image tokenizer' that breaks an original image into small visual pieces based on its learned set of patterns. Both picture patches and visual tokens are used to view each image during pre-training. After that, some of these picture patches are randomly masked with a unique mask embedding.

Before applying these transformer models to classify emotions, data preprocessing techniques such as random-sized crop, normalization, and resizing were applied to the Mel- spectrogram. These models were then trained with a batch size of 32 and 20 epochs.

To evaluate the performance of the models discussed in this study, accuracy, precision, recall, and F1-score were calculated with 5-fold stratified cross-validation in order to ensure balanced and reliable assessment across the distributions of the classes.

E. Classification Metrics

Classification metrics play a major role in this study. To evaluate the performance of CNN, LSTM, hybrid CNN and LSTM, ViT, and BEiT, accuracy, precision, recall, and F1- score were calculated using Equations (2), (3), (4), and (5), respectively.

$$\text{Accuracy} = (\text{True Positive} + \text{True Negative}) / (\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}) \quad (2)$$

$$\text{Precision} = (\text{True Positive}) / (\text{True Positive} + \text{False Positive}) \quad (3)$$

$$\text{Recall} = (\text{True Positive}) / (\text{True Positive} + \text{False Negative}) \quad (4)$$

$$\text{F1-score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (5)$$

After calculating the evaluation metrics, the model that achieved the highest accuracy, precision, recall, and F1-score was selected as the best-performing model, and to explain the output of this model, Grad-CAM, which is an explainable artificial intelligence (XAI) technique, was used.

F. The Explainable AI Technique - Grad-CAM

The complex artificial intelligence (AI) models are being applied in many industries. As a result, XAI is crucial to evaluate the predictions of those AI models, especially in fields like healthcare and finance. The main aim of these XAI models is to make the model's decision-making process transparent, which helps build trust in the model's predictions. In healthcare, explainability will increase trust of clinicians towards the predictions made by the AI model and can improve the security of patients. CAM (Class Activation Mapping) is a one XAI technique [27]. This used a global average pooling layer to replace a fully connected layer. This method results in a heatmap of an image for a specific class. This heatmap could explain how the CNN categorized the image as a particular class. CAM cannot generate a heatmap using

intermediate layers and is not compatible with transfer learning models. To overcome these limitations, Grad-CAM was introduced. Unlike CAM, Grad-CAM does not require a global average pooling layer but operates on the gradients of the target class score concerning the feature maps. These gradients are averaged globally to obtain importance weights, which are used similarly to CAM to compute a weighted sum over the feature maps. Lastly, a ReLU activation is applied to focus the visualization on the most effective positive regions, and thus Grad-CAM becomes a more generalizable and applicable tool for model interpretability [28].

IV. RESULTS AND DISCUSSION

This study used the TESS, which contained 28000 audio recordings created by two actresses. Firstly, acoustic features were extracted from the dataset, and three different classification methods, including CNN, LSTM, and a hybrid of CNN and LSTM, were applied. After that, evaluation metrics were calculated for each of these models, and Table 1 depicts the accuracies achieved by these three models.

Table 1. The accuracies of the CNN, LSTM, and the hybrid of CNN and LSTM models

Model	Layers used	Learning rate	Accuracy
CNN	Conv1D layer Max pooling layer Flatten layer Dropout layer Dense layer	0.001	95.37%
LSTM	LSTM layer Dropout layer Dense layer	0.001	98.57%
A hybrid of CNN and LSTM	Conv1D layer Max pooling layer Dropout layer LSTM layer	0.001	99.01%

According to Table 1, it is clear that the hybrid of the CNN and LSTM model performs better than the CNN model and the LSTM model, with an accuracy of 99.01%. However, the LSTM model achieved an accuracy of 98.57%, which is slightly lower than that of the hybrid model. In addition, to evaluate the performance of these three models in detail, precision, recall, and F1-score were calculated and are presented in Table 2.

When considering Table 2, it is observed that the hybrid model achieved the highest performance across three evaluation metrics, with a precision of 99.30%, a recall, and an F1-score of 99.29%. This highlights the hybrid model's accuracy and robustness in classification when compared to two individual CNN and LSTM models.

Table 2. The precision, recall, and F1-score of the CNN, LSTM, and the hybrid of CNN and LSTM models

Model	Precision	Recall	F1-score
CNN model	95.30%	95.30%	95.37%
LSTM model	98.55%	98.56%	98.57%
CNN+LSTM model	99.30%	99.29%	99.29%

However, to compare the performance of these classification models, two major classical image transformer models were applied to the Mel spectrograms obtained from the TESS dataset, and the accuracies calculated for these models are demonstrated in Table 3.

Table 3. The accuracies of the ViT and BEiT models

Model	Learning rate	Accuracy
ViT	0.0001	98%
BEiT	0.001	98.3%

When considering Table 3, it is clear that the BEiT model achieved the highest accuracy of 98.3% when compared to that of the ViT model. In addition to accuracy, precision, recall, and F1-score were calculated to provide a more comprehensive model evaluation. The calculated metrics are depicted in Table 4.

Table 4. The precision, recall, and F1-score of the ViT and BEiT models

Model	Precision	Recall	F1-score
ViT Model	98.00%	98.00%	98.00%
BEiT model	98.31%	98.30%	98.30%

As demonstrated in Table 4, the BEiT model outperforms the ViT model across all metrics by achieving a precision of 98.31%, a recall of 98.30%, and a F1-score of 98.30%. This indicates that the BEiT model performs well compared the ViT model in classification tasks. However, it is evident from Tables 1, 2, 3 and 4 that the CNN and LSTM hybrid model performs better with the dataset by obtaining higher accuracy as well as notable precision, recall, F1-score than the other suggested models. This indicates that this proposed hybrid model has a higher classification power when compared to both the novel and classical models. In order to provide additional evidence for hybrid model's robustness in effectively classifying emotion categories, classification report as well as the confusion matrix were obtained as in Figures 3, and 4 respectively.

According to the classification report and confusion matrix shown in Figures 3, and 4 respectively, it is clear that the hybrid model performs well with minimal misclassifications as across all classes demonstrating its

strong and balanced performance in emotion recognitions. Despite the model's higher values in evaluation metrics, it is essential to examine the accuracy and loss curves of this hybrid model. Therefore, the accuracy curve and the loss curve obtained for the hybrid model are depicted in Figures 5 and 6, respectively.

Classification Report:				
	precision	recall	f1-score	support
0	1.00	0.98	0.99	82
1	0.99	1.00	0.99	80
2	0.99	1.00	0.99	80
3	1.00	0.99	0.99	80
4	0.99	1.00	0.99	81
5	1.00	0.99	0.99	80
6	0.99	1.00	0.99	80
accuracy			0.99	563
macro avg	0.99	0.99	0.99	563
weighted avg	0.99	0.99	0.99	563
Weighted Precision: 0.9930				
Weighted Recall: 0.9929				
Weighted F1-Score: 0.9929				

Figure 3. Classification report of the hybrid of the CNN and LSTM model

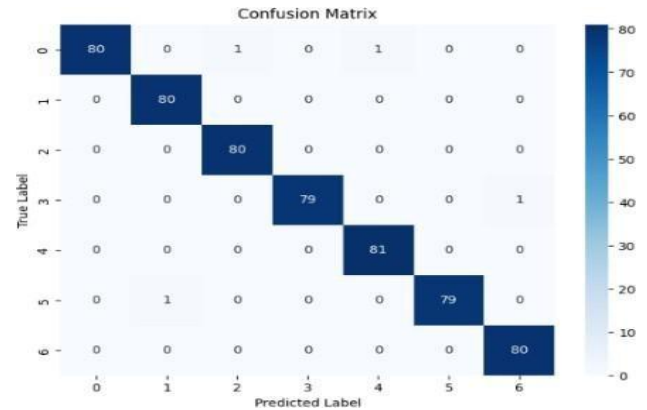


Figure 4. Confusion matrix of the hybrid of the CNN and LSTM model

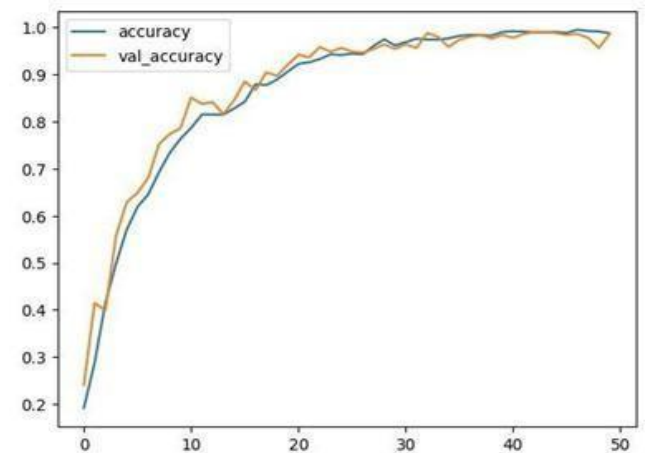


Figure 5. Accuracy curve of the hybrid of the CNN and LSTM model

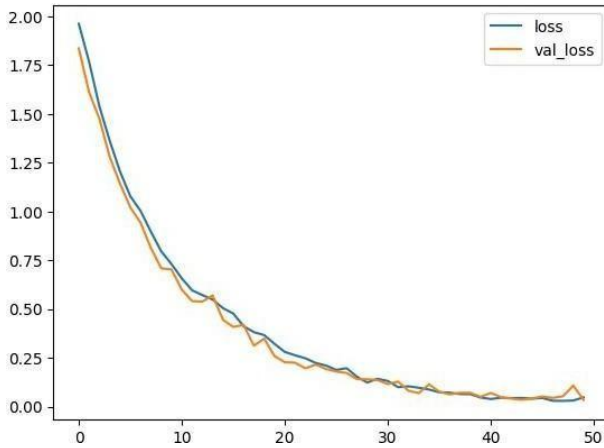


Figure 6. Loss curve of the hybrid of the CNN and LSTM model

According to Figures 5 and 6, it is clear that the suggested model learns effectively over time. Therefore, this study concludes that the hybrid CNN and LSTM model, under the given conditions, is more suitable for speech emotion recognition. However, to observe how this proposed model made its final predictions, Figures 7,8, 9, 10,11,12, and 13 were generated to show how Grad CAM evaluation was performed for classes angry, fear, disgust, happy, sad, surprise, and neutral predictions in the test dataset, respectively. To explain briefly, additional metrics are shown in each Grad-CAM figure.

The Grad-CAM visualization for the "angry" emotion class, which is represented in Figure 7, highlights significant time- frequency regions in the Mel-spectrogram that contributed mainly to the classification decision of the model. The red and yellow high-activation regions indicate where the model was focusing, which likely corresponds with speech features typical in anger, such as high energy and pitch. The model labelled "angry" with 77.49% confidence, concentrating on a small but crucial region (1.74% of the spectrogram) between 0.22–2.59 seconds and 367.0–6899.1 Hz. Masking this region resulted in a very slight decrease in confidence (0.54%), confirming its importance. The model's attention to these specific features allows for the interpretation and verification of its emotion classification decision.

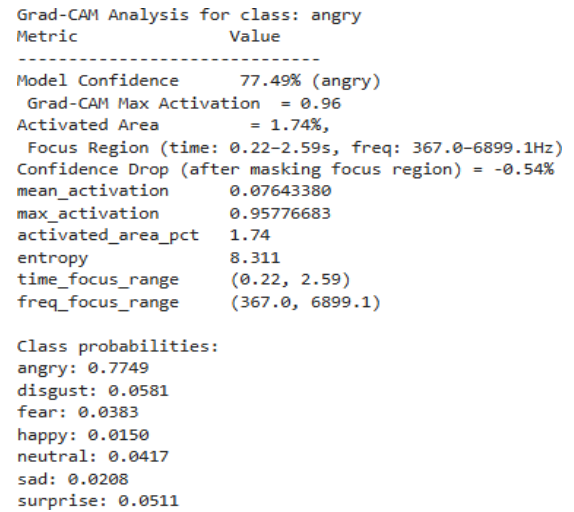
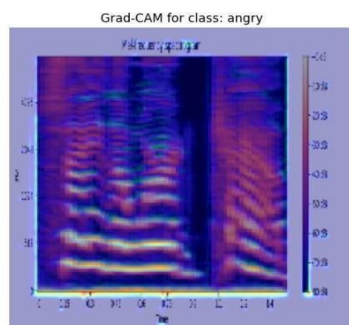
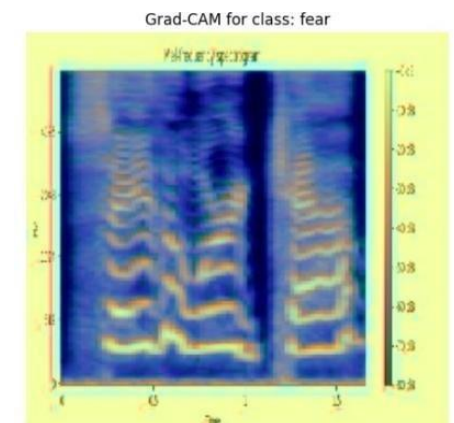


Figure 7. Grad-CAM output for angry class test images

Figure 8 shows the Grad-CAM representation for the fear class. The Grad-CAM visualization of the fear class highlights significant time-frequency regions in the Mel-spectrogram that contributed significantly to the classification decision of the model. The amount of yellow and red colour can be seen throughout the graph, and this is as a result of high-activation regions indicating where the model was focusing, which likely corresponds with speech features typical in fear. The model labelled "fear" with 54.78% confidence, concentrating on a small but crucial region (42.20% of the spectrogram) between 0.0–2.97 seconds and 0.0–7926.6 Hz. Masking this region resulted in a drop in confidence (19.66%), confirming its importance.

In addition, Figures 8, 12, and 13 show more yellow in colour since the model is more confident that those regions contributed to its final decision. The model confidence levels are very high in these Grad-CAM figures.



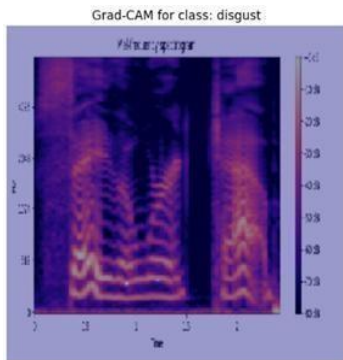
Grad-CAM Analysis for class: fear

Metric	Value
Model Confidence	54.73% (fear)
Grad-CAM Max Activation	= 0.98
Activated Area	= 44.20%,
Focus Region (time: 0.0-2.97s, freq: 0.0-7926.6Hz)	
Confidence Drop (after masking focus region)	= 19.66%
mean_activation	0.44908139
max_activation	0.98126942
activated_area_pct	44.20
entropy	9.310
time_focus_range	(0.0, 2.97)
freq_focus_range	(0.0, 7926.6)

Class probabilities:

angry: 0.0625
disgust: 0.0671
fear: 0.5473
happy: 0.1223
neutral: 0.0554
sad: 0.0178
surprise: 0.1277

Figure 8. Grad-CAM output for fear class test images



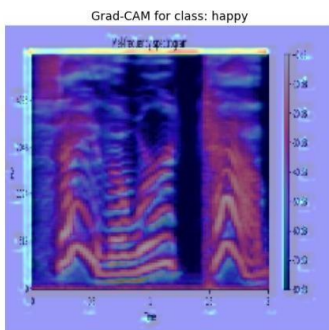
Grad-CAM Analysis for class: disgust

Metric	Value
Model Confidence	52.23% (disgust)
Grad-CAM Max Activation	= 0.53
Activated Area	= 0.01%,
Focus Region (time: 1.05-1.05s, freq: 6091.7-6091.7Hz)	
Confidence Drop (after masking focus region)	= 0.00%
mean_activation	0.00008228
max_activation	0.52723080
activated_area_pct	0.01
entropy	1.018
time_focus_range	(1.05, 1.05)
freq_focus_range	(6091.7, 6091.7)

Class probabilities:

angry: 0.0208
disgust: 0.5223
fear: 0.0502
happy: 0.1162
neutral: 0.0984
sad: 0.0924
surprise: 0.0996

Figure 9. Grad-CAM output for disgust class test images



Grad-CAM Analysis for class: happy

Metric	Value
Model Confidence	71.81% (happy)
Grad-CAM Max Activation	= 0.94
Activated Area	= 1.62%,
Focus Region (time: 0.17-2.72s, freq: 293.6-7559.6Hz)	
Confidence Drop (after masking focus region)	= -0.99%
mean_activation	0.12068452
max_activation	0.93679237
activated_area_pct	1.62
entropy	8.858
time_focus_range	(0.17, 2.72)
freq_focus_range	(293.6, 7559.6)

Class probabilities:

angry: 0.0179
disgust: 0.0326
fear: 0.0432
happy: 0.7181
neutral: 0.0663
sad: 0.0266
surprise: 0.0953

Figure 10. Grad-CAM output for happy class test images

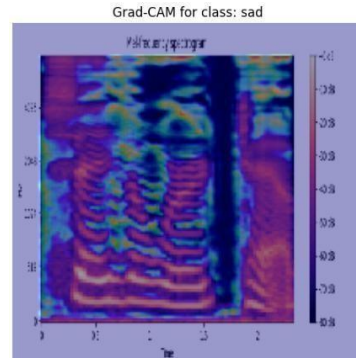


Figure 11. Grad-CAM output for sad class test images

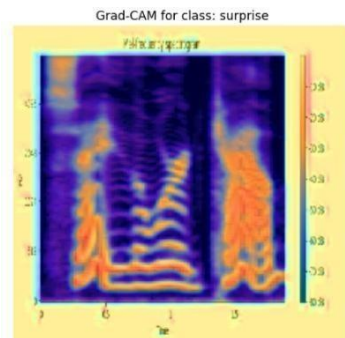
Grad-CAM Analysis for class: sad

Metric	Value
Model Confidence	71.87% (sad)
Grad-CAM Max Activation	= 0.92
Activated Area	= 3.64%,
Focus Region (time: 0.25-2.37s, freq: 733.9-6678.9Hz)	
Confidence Drop (after masking focus region)	= 3.09%
mean_activation	0.08298303
max_activation	0.91536552
activated_area_pct	3.64
entropy	7.960
time_focus_range	(0.25, 2.37)
freq_focus_range	(733.9, 6678.9)

Class probabilities:

angry: 0.0119
disgust: 0.0536
fear: 0.0152
happy: 0.0407
neutral: 0.1308
sad: 0.7187
surprise: 0.0291

Figure 11. Grad-CAM output for sad class test images



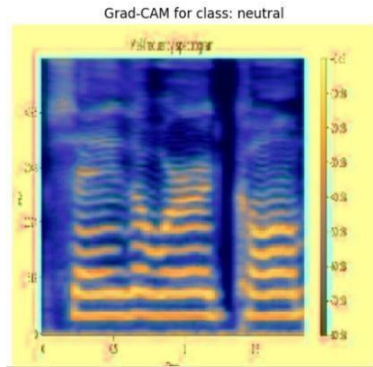
```

Grad-CAM Analysis for class: surprise
Metric      Value
-----
Model Confidence      79.13% (surprise)
Grad-CAM Max Activation = 0.98
Activated Area      = 51.18%,
Focus Region (time: 0.0-2.97s, freq: 0.0-7926.6Hz)
Confidence Drop (after masking focus region) = 55.17%
mean_activation      0.41372573
max_activation      0.98044980
activated_area_pct    51.18
entropy      9.080
time_focus_range      (0.0, 2.97)
freq_focus_range      (0.0, 7926.6)

Class probabilities:
angry: 0.0091
disgust: 0.0315
fear: 0.0158
happy: 0.0541
neutral: 0.0767
sad: 0.0214
surprise: 0.7913

```

Figure 12. Grad-CAM output for surprise class test images



```

Grad-CAM Analysis for class: neutral
Metric      Value
-----
Model Confidence      91.82% (neutral)
Grad-CAM Max Activation = 0.93
Activated Area      = 51.21%,
Focus Region (time: 0.0-2.97s, freq: 0.0-7926.6Hz)
Confidence Drop (after masking focus region) = 74.96%
mean_activation      0.46167940
max_activation      0.93491471
activated_area_pct    51.21
entropy      9.281
time_focus_range      (0.0, 2.97)
freq_focus_range      (0.0, 7926.6)

Class probabilities:
angry: 0.0064
disgust: 0.0086
fear: 0.0047
happy: 0.0113
neutral: 0.9182
sad: 0.0213
surprise: 0.0296

```

Figure 13. Grad-CAM output for neutral class test images

Therefore, we can conclude that the model predictions are correct and that we can use the Grad-CAM to evaluate the proposed model. These methods can reduce the black box nature of AI model predictions and can increase trust among AI models, clinicians, and researchers.

However, despite the promising results, this study faces several limitations. The dataset, which was used in this

study, contained audio files generated by two actresses. As a result, these models fail to capture the gender variations in the speech emotion recognition process. Furthermore, by limiting it to two actresses of two different ages, it only identifies speech emotions at these ages. Therefore, hidden patterns at different age limits may not be included in the trained model. In addition, this study is restricted to seven emotions, which may not fully capture the complexity and subtlety of human emotional expression, potentially restricting the model's ability to generalize to real-world emotional variability. Furthermore, this study compared the performance of the proposed model with two other classification methods and classical image transformers, limiting the scope of evaluation. However, this study could be used as a benchmark for future research, and these limitations could be addressed in further studies using not only human audio but also sounds related to birds, animals, and environmental contexts, etc. Moreover, enhancing this study by including audio from different age groups, genders, and cultural backgrounds would provide more robust results. Furthermore, to improve the generalizability as well as interpretability, future research could improve this study using diverse datasets and evaluating the model's performance across various datasets. In addition, this study opens the pathway for implementing and evaluating other neural network models to recognize speech emotions. This would lead to more generalized results and conclusions, thereby contributing to speech emotion recognition systems.

V. CONCLUSION

This study implemented CNN, LSTM, and a CNN-LSTM hybrid model using the TESS dataset. MFCC, chroma, Mel- scaled spectrogram, and contrast feature extraction techniques were used to extract the acoustic features prior to implementing the above-mentioned models. The CNN and LSTM hybrid model achieved a remarkable accuracy of 99.01%, while the CNN and LSTM models separately achieved a better performance of 95.37% and 98.57%, respectively. To assess the performance of these classification models, two classical image transformer models, including ViT and BEiT, were employed on the Mel-spectrogram of audio files from the TESS dataset, and it was identified that the ViT model acquired an accuracy of 98%, while the BEiT model acquired a better accuracy of 98.3%. However, these two transformer models were unable to outperform the hybrid model in terms of accuracy. Moreover, to evaluate the output of this model thoroughly, the Grad-CAM, which is a novel explainable AI technique, was used in this study. While the TESS dataset is a widely used, well-annotated benchmark for SER tasks, it was collected from only two female speakers aged 26 and 64. Thus, our findings might not generalize well across genders, ages, and speaking styles. This reduced demographic coverage might bring bias or reduce robustness when applied to large populations. Therefore, in future work, we plan to include other datasets such as Ryerson Audio-Visual Database of Emotional Speech and

Song (RAVDESS) or Crowd-Sourced Emotional Multimodal Actors Dataset (CREMA-D) that offer higher speaker variability. Overall, despite these limitations, the current work is a strong baseline and demonstrates the potential of the proposed hybrid CNN-LSTM model when applied to well-annotated emotional speech data.

REFERENCES

- [1] W. Zheng, W. Zheng, and Y. Zong, "Multi-scale discrepancy adversarial network for crosscorpus speech emotion recognition," *Virtual Reality & Intelligent Hardware*, vol. 3, no. 1, pp. 65–75, Feb. 2021, doi: <https://doi.org/10.1016/j.vrih.2020.11.006>.
- [2] G. A. Koduru, H. B. Valiveti, and A. K. Budati, "Feature extraction algorithms to improve the speech emotion recognition rate," *International Journal of Speech Technology*, vol. 23, no. 1, pp. 45–55, Jan. 2020, doi: <https://doi.org/10.1007/s10772-020-09672-4>.
- [3] J. H. L. Hansen and D. A. Cairns, "ICARUS: Source generator based real-time recognition of speech in noisy stressful and Lombard effect environments ☆," *Speech Communication*, vol. 16, no. 4, pp. 391–422, Jun. 1995, doi: [https://doi.org/10.1016/0167-6393\(95\)00007-b](https://doi.org/10.1016/0167-6393(95)00007-b).
- [4] C. Spencer et al., "A Comparison of Unimodal and Multimodal Measurements of Driver Stress in Real-World Driving Conditions," *PsyArXiv (OSF Preprints)*, Jun. 2020, doi: <https://doi.org/10.31234/osf.io/en5r3>.
- [5] B. Schuller, G. Rigoll, and M. Lang, "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture," *IEEE Xplore*, May 01, 2004, doi: <https://ieeexplore.ieee.org/document/1326051> (accessed Feb. 26, 2021).
- [6] D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, and D. M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk," *IEEE transactions on bio-medical engineering*, vol. 47, no. 7, pp. 829–837, Jul. 2000, doi: <https://doi.org/10.1109/10.846676>.
- [7] "M. Young, 'The Technical Writers Handbook,' Mill Valley, CA University Science, 1989. - References - Scientific Research Publishing," *Scirp.org*, 2021. <https://www.scirp.org/reference/referencespapers?referenceid=998786> (accessed Sep. 01, 2024).
- [8] "Speech and Multimedia Transmission Quality (STQ); Requirements for Emotion Detectors used for Telecommunication Measurement Applications; Detectors for written text and spoken speech TECHNICAL SPECIFICATION." Accessed: Sep. 01, 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/103200_103299/103296/01.01.01_60/ts_103296v010101p.pdf
- [9] S. G. Koolagudi and K. S. Rao, "Emotion recognition from speech: a review," *International Journal of Speech Technology*, vol. 15, no. 2, pp. 99–117, Jan. 2012, doi: <https://doi.org/10.1007/s10772-011-9125-1>.
- [10] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, Mar. 2011, doi: <https://doi.org/10.1016/j.patcog.2010.09.020>.
- [11] M. Ren, W. Nie, A. Liu, and Y. Su, "Multi-modal Correlated Network for emotion recognition in speech," *Visual Informatics*, vol. 3, no. 3, pp. 150–155, Sep. 2019, doi: <https://doi.org/10.1016/j.visinf.2019.10.003>.
- [12] "Indexof/class/archive/cs/cs224n/cs224n.1214/reports," *Stanford.edu*, 2021. <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1214/reports/> (accessed Sep. 01, 2024).
- [13] M. M. Rezapour Mashhadi and K. Osei-Bonsu, "Speech emotion recognition using machine learning techniques: Feature extraction and comparison of convolutional neural network and random forest," *PloS one*, vol. 18, no. 11, p. e0291500, 2023, doi: <https://doi.org/10.1371/journal.pone.0291500>.
- [14] N. P. Tigga and S. Garg, "Speech Emotion Recognition for multiclass classification using Hybrid CNN-LSTM," *International Journal of Microsystems and Iot*, vol. 1, pp. 9–17, 2023.
- [15] H. Qazi and B. N. Kaushik, "A hybrid technique using CNN+ LSTM for speech emotion recognition," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 5, pp. 1126–1130, 2020.
- [16] L. Kerkeni, Y. Serrestou, M. Mbarki, K. Raoof, M. A. Mahjoub, and C. Cleder, "Automatic Speech Emotion Recognition Using Machine Learning," *social media and Machine Learning*, Mar. 2019, doi: <https://doi.org/10.5772/intechopen.84856>.
- [17] H. S. Kumbhar and S. U. Bhandari, "Speech Emotion Recognition using MFCC features and LSTM network," Sep. 2019, doi: <https://doi.org/10.1109/iccubea47591.2019.9129067>

- [18] Y. Yu and Y.-J. Kim, "Attention-LSTM-Attention Model for Speech Emotion Recognition and Analysis of IEMOCAP Database," *Electronics*, vol. 9, no. 5, p. 713, Apr. 2020, doi: <https://doi.org/10.3390/electronics9050713>.
- [19] "Speech Emotion Recognition Using CNN-LSTM and Vision Transformer," *Dntb.gov.ua*, 2023. <https://ouci.dntb.gov.ua/en/works/7WQ2BrPl/> (accessed Sep. 01, 2024).
- [20] "Toronto emotional speech set (TESS)," *www.kaggle.com*. <https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess>
- [21] L. Toledo, A. Luiz, and J. Fiais, "A Deep Learning Approach for Speech Emotion Recognition Optimization Using Meta- Learning," *Electronics*, vol. 12, no. 23, pp. 4859–4859, Dec. 2023, doi: <https://doi.org/10.3390/electronics12234859>.
- [22] V. Vielzeuf, S. Pateux, and F. Jurie, "Temporal multimodal fusion for video emotion classification in the wild," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017, pp. 569–576.
- [23] S. Waldekar and G. Saha, "Wavelet Transform Based Mel- scaled Features for Acoustic Scene Classification.," in *INTERSPEECH*, 2018, vol. 2018, pp. 3323–3327.
- [24] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [25] "Baeldung on CS," *www.baeldung.com*, Mar. 19, 2021. <https://www.baeldung.com/cs/>.
- [26] H. Bao, L. Dong, S. Piao, and F. Wei, "Beit: Bert pre-training of image transformers," *arXiv preprint arXiv:2106.08254*, 2021.
- [27] Zhou, Bolei, et al. "Learning deep features for discriminative localization." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [28] Jeong, Seung-Min, et al. "Exploring Spectrogram-Based Audio Classification for Parkinson's Disease: A Study on Speech Classification and Qualitative Reliability Verification." *Sensors* 24.14 (2024): 4625.

AUTHOR BIOGRAPHY

H.M.L.S. Kumari

H.M.L.S. Kumari, an Instructor at the Computer Center, Faculty of Engineering, University of Peradeniya, completed a Bachelor of Science Honours degree in Computer Science at the Faculty of Applied Sciences, Vavuniya Campus, University of Jaffna. Her research interests include Deep Learning, Computer Vision, Artificial Intelligence, and Explainable AI, particularly in the health sector.

H.M.N.S. Kumari

H.M.N.S. Kumari is a volunteer researcher currently engaged in numerous projects. She completed her Master of Science degree in Computing Sciences, specializing in Statistical Data Analytics at Tampere University, Finland, and her Bachelor of Science degree in Statistics and Operations Research at the University of Peradeniya, Sri Lanka. Her research interests include Statistical Data Analytics, Bayesian Statistics, Machine Learning, Deep learning, Computer Vision, Explainable Artificial Intelligence, Signal Processing, and Public Health.

U.M.M.P.K. Nawarathne

U.M.M.P.K. Nawarathne, an Assistant Lecturer at the Faculty of Computing, Sri Lanka Institute of Information Technology, completed her Bachelor of Science Honors Degree in Information Technology, specializing in Data Science, in 2021. Her main research interests lie in the areas of Generative Artificial Intelligence and applied Data Science.

Journal website:

