

**Deep Learning Tool for Higher Learning Scholarship Award Decisions: A
Case of Higher Institutions of Learning**

By

Nyambura, Thomas Macharia

**Submitted in partial fulfilment of the requirements for the Degree of
Master of Science in Information Technology at Strathmore University**

School of Computing and Engineering Sciences

Strathmore University

Nairobi, Kenya

June, 2025

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

© No part of this thesis may be reproduced without the permission of the author and Strathmore University

Student's Name: Nyambura, Thomas Macharia

Sign:  Date: 29-05-2025

Approval

The thesis of Nyambura, Thomas Macharia was reviewed and approved for examination by the following:

Dr. Esther Khakata,
School of Computing & Engineering Sciences,
Strathmore University

Dr. Julius Butime,
Dean, School of Computing & Engineering Sciences,
Strathmore University

Prof. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

In today's highly competitive higher education environment, scholarship programs play a crucial role in attracting and retaining talented students, especially those from disadvantaged backgrounds. However, traditional scholarship award processes often rely on manual and subjective methods, which can lead to inefficiencies, biases, and inequitable distribution of financial aid. The lack of standardized criteria and the complexity of handling large volumes of applications further complicates the decision-making process. This study developed a deep learning-based tool to assist in higher education scholarship award decisions, utilizing convolutional neural network (CNN) to automate the selection process. Through a data-driven approach, the tool was designed to analyse student profiles across multiple dimensions, including academic performance, financial need, extracurricular involvement, and personal background. The deep learning model aims to provide more accurate, objective, and fair scholarship allocations by removing human biases and ensuring that all eligible students are assessed equitably. The developed tool was tested on a dataset of historical scholarship applications and awards, and the performance of the CNN model was evaluated based on accuracy, fairness, and efficiency compared to traditional manual methods. The model achieved an accuracy of 93% surpassing the performance of all the reviewed models at the end of fine-tuning. The findings of this study offer valuable insights into the feasibility and effectiveness of integrating AI-driven tools into scholarship management systems in higher education institutions. Ultimately, this study contributes to ongoing efforts to make scholarship distribution more equitable and aligned with institutional goals and societal needs.

Keywords: *Scholarship Allocation, Deep Learning, Convolutional neural networks (CNNs), machine learning, higher education, automated decision-making, data-driven scholarship distribution, AI in education*

Table of Contents

Declaration and Approval	ii
Abstract	iii
Table of Contents	iv
List of Figures	ix
List of Tables	xi
Abbreviations and Acronym	xii
Acknowledgements.....	xiii
Definition of Terms.....	xiv
Chapter 1: Introduction	1
1.1 Background.....	1
1.2 Problem Statement.....	3
1.3 Research Objectives.....	4
1.4 Research Questions.....	4
1.5 Justification.....	4
1.6 Scope and Limitations.....	5
Chapter 2: Literature Review.....	6
2.1 Introduction.....	6
2.2 Challenges in the Current Scholarship Allocation Processes	6
2.2.1 Inefficiency in Manual Processes	6
2.2.2 Bias and Inequity in Scholarship Allocation	7
2.2.3 Lack of Standardization in Scholarship Criteria.....	8
2.2.4 Difficulty in Measuring Holistic Student Potential	8
2.3 Theoretical Literature.....	9
2.3.1 Human Capital Theory.....	9
2.3.2 Meritocracy Theory	9
2.3.3 Equity Theory	10
2.4 Empirical Literature	11
2.5 Existing Machine Learning Models and Algorithms for Scholarship Award Decisions.	15
2.5.1 Models.....	15

2.5.1.1	Ensemble Model	15
2.5.1.2	Random Forest Model	16
2.5.1.3	Support Vector Model (SVM).....	18
2.5.1.4	Multilayer Perceptron	19
2.5.2	Frameworks.....	20
2.5.2.1	TensorFlow	20
2.5.2.2	Scikit-Learn.....	21
2.5.2.2	Keras	21
2.6	Architectures and Designs	21
2.6.1	Data Mining Architecture	22
2.6.2	Naive Bayes and Random Forest Architecture.....	22
2.6.3	Holistic Architecture.....	23
2.6.4	Student Scholarship prediction Architecture	24
2.7	Algorithms	26
2.7.1	ARIMA algorithm.....	26
2.7.2	Genetic Algorithms.....	26
2.7.3	Support Vector Machine (SVM).....	27
2.7.4	Artificial Neuron Network (ANN).....	28
2.7.5	Naïve Bayes	29
2.7.6	K-Nearest Neighbors (KNN).....	30
2.8	Gaps in Literature	31
2.9	Conceptual Model.....	32
Chapter 3: Research Methodology.....		34
3.1	Introduction.....	34
3.2	Research Design and Philosophy.....	34
3.3	Population and Sampling	35
3.3.1	Population	35
3.3.2	Sampling	35
3.4	Data Collection Methods and Analysis.....	35
3.4.1	Data Collection	35
3.4.2	Data Analysis	36

3.5 Research Quality and Reliability	36
3.6 Systems Development Methodology	37
3.6.1 Planning	37
3.6.2 Design	38
3.6.3 Development	38
3.6.4 Testing.....	39
3.6.5 Deployment.....	39
3.6.6 Review and Feedback	39
3.6.7 Maintenance and Updates	39
3.7 Utilization and Dissemination of Research Results.....	40
3.8 Ethical Considerations and Issues.....	40
Chapter 4: System Analysis and Design.....	41
4.1 Introduction.....	41
4.2 Requirement Specifications	41
4.2.1 Functional Requirements	41
4.2.2 Non-Functional Requirements.....	42
4.3 System Architecture.....	42
4.4 System Design	43
4.4.1 Use Case Diagram.....	43
3.8.1.1 4.4.1.1 Detailed Use Case Descriptions	44
4.4.2 Class Diagram.....	45
4.4.3 Sequence Diagram	45
4.4.4 Database Schema	46
4.5 Wireframes.....	47
4.5.1 Home Page Wireframe.....	47
4.5.2 Login Wireframe.....	48
4.5.3 Scholarship Application Wireframe	49
4.5.4 Scholarship Decision Wireframe	50
4.5.5 Register Wireframe.....	51
Chapter 5: System Implementation and Testing.....	53

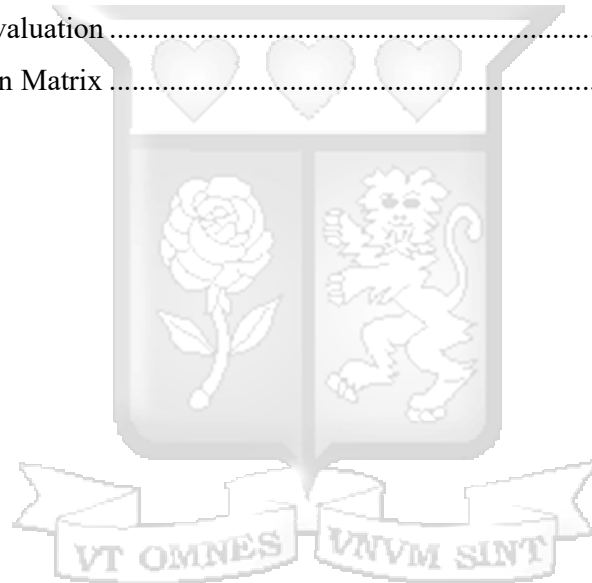
5.1 Introduction.....	53
5.2 Model Components.....	53
5.2.1 Overall CNN Architecture.....	53
5.2.2 Convolutional Layers.....	54
5.2.3 Pooling Layers.....	55
5.2.4 Dropout Layers.....	55
5.2.5 Flattening Layer.....	55
5.2.6 Dense Layers.....	56
5.2.7 Model Compilation.....	56
5.3 Scholarship Award Decision Tool.....	56
5.3.1 Landing Page.....	57
5.3.2 Scholarship Application Page.....	57
5.3.3 Application Results Page.....	58
5.4 System Implementation.....	59
5.4.1 Development Environment.....	59
5.4.2 Scholarship Data Collection.....	60
5.4.3 Data Pre-processing.....	60
5.4.3.1 Loading the Dataset.....	60
5.4.3.2 Handling Missing Values.....	60
5.4.3.3 Dropping Irrelevant Columns.....	61
5.4.4.4 Encoding Categorical Variables.....	61
5.4.4.5 Separating Features and Target Variable.....	61
5.4.4.6 Scaling Numerical Features.....	62
5.4.4.8 Splitting Data.....	62
5.4.4 Training Model.....	63
5.4.4.1 Model Definition.....	63
5.4.4.2 Model Compilation.....	63
5.4.4.3 Training Process.....	64
5.4.4.5 Fine-Tuning.....	64
5.4.4.6 Model Evaluation.....	64
5.4.5 Flask API.....	64

5.5 System Testing.....	65
5.5.1 Test on Model Accuracy	65
5.5.2 Functional Testing	67
5.5.3 Scalability Testing	67
Chapter 6: Discussions.....	68
6.1 Introduction.....	68
6.2 Review of Study Objectives.....	68
6.2.1 Challenges in the Current Scholarship Allocation Processes	68
6.2.2 Existing Machine Learning Models and Algorithms Used For Scholarship Award Decisions.....	69
6.2.3 Deep Learning-Based Scholarship Award Decision Tool Using CNN	70
6.2.4 System Testing.....	70
6.4.1.1 Model Accuracy.....	70
6.4.4.2 Functional Testing	71
6.4.4.3 Scalability Testing	71
Chapter 7: Conclusion and Recommendation.....	72
7.1 Conclusion	72
7.2 Recommendations.....	73
7.3 Future work.....	73
7.4 Limitations	74
7.5 Research Contribution	74
References.....	76
Appendices.....	83
Appendix A: Similarity Report.....	83
Appendix B: Ethical Clearance Certification	91

List of Figures

Figure 2.1 Architecture of Ensemble Model	16
Figure 2.2 Random Forest Model Architecture	17
Figure 2.3 SVM Model Architecture.....	19
Figure 2.4 MLP Model Architecture	20
Figure 2.5 Data Mining Architecture.....	22
Figure 2.6 Naive Bayes and Random Forest Architecture.	23
Figure 2.7 Holistic Architecture	24
Figure 2.8 Holistic Architecture	25
Figure 2.9 Structure of ANN.	29
Figure 2.10 Conceptual Model	33
Figure 3.1 Agile Methodology.....	37
Figure 4.1: System Architecture	43
Figure 4.2: Use Case Diagram	44
Figure 4.3 Class Diagram	45
Figure 4.4: Sequence Diagram.....	46
Figure 4.5: Database Schema.....	47
Figure 4.6: Home Page Wireframe	48
Figure 4.7: Login Wireframe	49
Figure 4.8: Scholarship Application Form.....	50
Figure 4.9: Scholarship Decision Wireframe.....	51
Figure 4.10: Register Wireframe	52
Figure 5.1: CNN Architecture.....	54
Figure 5.2: Convolution Layer.....	54
Figure 5.3 Convolution Layers	55
Figure 5.4 Pooling Layers.....	55
Figure 5.5 Dropout Layers	55
Figure 5.6 Flattening Layers	56
Figure 5.7 First Dense Layer	56
Figure 5.8 Second Dense Layer	56
Figure 5.9 Model Compilation.....	56
Figure 5.10: Landing Page.....	57
Figure 5.11: Scholarship Application Page	58

Figure 5.12 Results Page.....	59
Figure 5.13: Load CSV file.....	60
Figure 5.14 Handling Missing Values	60
Figure 5.15 Dropping Irrelevant Columns.....	61
Figure 5.16 Encoding Categorical Variables	61
Figure 5.17: Separating Features and Target Variables.....	62
Figure 5.18: Scaling Numerical Columns.....	62
Figure 5.19: Reshaping Data for CNN Input.....	62
Figure 5.20: Splitting Data.....	62
Figure 5.21: Model Training.....	64
Figure 5.22: API.....	65
Figure 5.23: Model Evaluation.....	66
Figure 5.24: Confusion Matrix	66



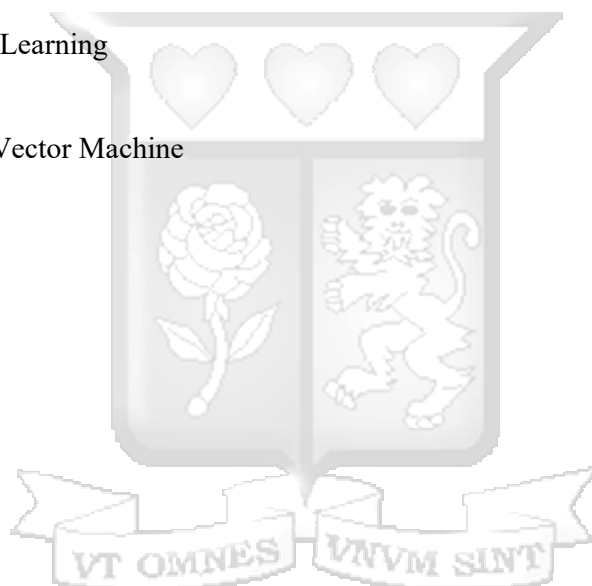
List of Tables

Table 2.1 Summary of the Empirical Studies	14
Table 2.2 Summary of Algorithms.	31
Table 4.1: Description of use cases.....	44



Abbreviations and Acronym

AI	Artificial Intelligence
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DL	Deep Learning
KNN	K-Nearest Neighbours
ML	Machine Learning
SVM	Support Vector Machine



Acknowledgements

I would like to express my deepest gratitude to all those who supported me throughout the completion of this thesis. I am especially thankful to my supervisor Dr. Esther Khakata, for her guidance, encouragement, and insightful feedback that significantly shaped the direction of this work. Special thanks to my family and friends, whose unwavering support and understanding have been a constant source of strength. Lastly, I am grateful to Strathmore University and my colleagues for providing a stimulating academic environment that has been crucial to my research journey.



Definition of Terms

Artificial Intelligence

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines designed to perform tasks that typically require human cognition (Sheikh et al., 2023).

Convolutional Neural Networks

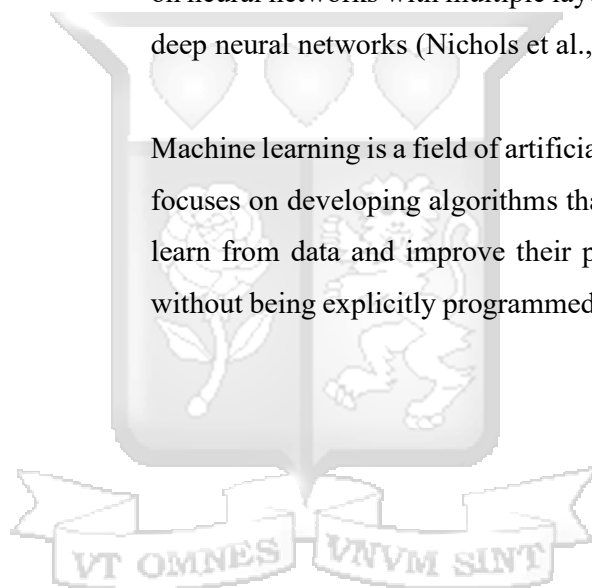
CNNs are a specialized type of deep learning neural network designed specifically for processing structured grid-like data, such as images (Yamashita et al., 2018).

Deep Learning

Deep learning is a subset of machine learning that focuses on neural networks with multiple layers, often referred to as deep neural networks (Nichols et al., 2018).

Machine Learning

Machine learning is a field of artificial intelligence (AI) that focuses on developing algorithms that enable computers to learn from data and improve their performance over time without being explicitly programmed (Nichols et al., 2018).



Chapter 1: Introduction

1.1 Background

In today's competitive landscape of higher education, scholarships are pivotal in enabling access to academic opportunities for students who might otherwise be unable to afford it. Scholarships serve as a key mechanism for promoting educational equity by offering financial aid to students who demonstrate academic merit or financial need (Budiyanto & Suryadi, 2023). However, the traditional methods of awarding scholarships often face significant challenges, particularly when it comes to fairness, transparency, and efficiency (Aulck et al., 2020b). As the demand for scholarships continues to grow, higher education institutions must explore innovative approaches to streamline the allocation process and ensure that scholarships reach the most deserving candidates.

Historically, the process of awarding scholarships has relied heavily on manual reviews and subjective decision-making by scholarship committees. Institutions typically evaluate applicants based on criteria such as academic performance, extracurricular activities, financial need, and letters of recommendation. However, this manual process is time-consuming and prone to biases, as it often depends on the discretion of the reviewing committee. Subjective factors such as personal biases, human error, and inconsistent evaluation standards can impact the fairness and accuracy of the scholarship allocation process, leading to inequitable outcomes where deserving students are overlooked (Irfan et al., 2024).

The scholarship allocation process is becoming increasingly complex with institutions having to manage a large volume of applications, each with a variety of data points to consider (Chaudhari et al., 2022). The growing demand for financial aid, especially in private and public universities, adds pressure on scholarship committees to make timely and accurate decisions. The higher the number of applications, the more difficult it becomes to manually assess each candidate's eligibility in an objective manner (Jaisri et al., 2024). This can lead to delays in decision-making, which in turn affects students' ability to plan their academic future. Furthermore, the subjective nature of manual evaluations means that decisions may be influenced by factors that are not directly related to a student's qualifications or financial need, further compromising the equity of the process (Gani et al., 2024).

The inherent limitations of traditional scholarship allocation processes highlight the need for a more systematic, data-driven approach. In recent years, technological advancements,

particularly in the field of artificial intelligence (AI) and machine learning (ML), have opened new avenues for transforming the scholarship award process (Mohd et al., 2024). AI-driven models, such as deep learning algorithms, can analyze large datasets with high accuracy and precision, offering a more efficient way to assess candidates based on multiple criteria. These models use data mining techniques to identify patterns in student profiles and predict scholarship eligibility more accurately than manual methods (Delima, 2019). By automating parts of the decision-making process, AI can help institutions reduce the time spent on evaluations, minimize human error, and ensure a more objective and fair allocation of scholarships.

One promising application of AI in scholarship allocation is the use of deep learning, a subset of machine learning that focuses on training algorithms to recognize patterns and make predictions from complex datasets (Sengupta et al., 2022). Specifically, convolutional neural networks (CNNs), a type of deep learning algorithm, have shown great potential in areas such as image recognition, healthcare diagnostics, and natural language processing. These algorithms can be adapted to the scholarship award process by analyzing various student data points, such as academic records, extracurricular activities, financial information, and demographic factors, to predict which candidates are most likely to benefit from financial aid. CNNs are capable of handling vast amounts of data with a high degree of accuracy, making them an ideal tool for automating the scholarship decision-making process (Chaudhari et al., 2022).

The integration of AI-driven scholarship systems has the potential to revolutionize the way scholarships are allocated by making the process more efficient, transparent, and fair. By leveraging deep learning algorithms, institutions can analyze a broader range of student data more comprehensively and consistently. This allows for a more holistic evaluation of candidates, considering factors that may be overlooked in traditional evaluation methods, such as socioeconomic background, future potential, and alignment with institutional goals (Irfan et al., 2024). Furthermore, AI models can be trained to minimize bias by removing subjective human judgment from the equation, thereby promoting a more equitable distribution of scholarships.

Another advantage of using deep learning models for scholarship allocation is the scalability of the approach. As institutions handle larger and more diverse applicant pools, the ability of AI systems to process thousands of applications quickly and accurately becomes increasingly

valuable. Unlike manual processes, which can become overwhelmed by large datasets, AI models can be scaled to accommodate a growing number of applications without sacrificing accuracy or efficiency (Gani et al., 2024). This scalability ensures that institutions can continue to offer financial aid to deserving students even as the demand for scholarships increases.

The potential benefits of AI in improving the scholarship allocation process are significant. By providing a data-driven approach to decision-making, deep learning models can help institutions allocate scholarships more effectively, ensuring that financial aid reaches the students who need it most. This approach not only improves the fairness and efficiency of the process but also aligns scholarship distribution with broader institutional goals, such as promoting diversity, inclusion, and academic excellence. As higher education institutions continue to adopt AI technologies, the scholarship allocation process will likely become more streamlined, objective, and equitable, ultimately benefiting both students and educational institutions (Chaudhari et al., 2022).

This study developed a deep learning scholarship allocation which will offer a transformative opportunity to address the limitations of traditional processes. The developed tool will help institutions ensure that scholarships are awarded based on objective criteria, minimizing bias and promoting fairness in higher education. As institutions continue to embrace technological innovations, the development of AI-driven scholarship tool will play a critical role in shaping the future of financial aid distribution in higher education.

1.2 Problem Statement

Traditional scholarship allocation methods in higher education often rely on manual review processes and subjective decision-making, leading to inefficiencies, biases, and suboptimal outcomes. As student data becomes more complex and diverse, it becomes increasingly difficult for administrators to fairly and efficiently match deserving students with scholarship opportunities. Moreover, the growing number of scholarship applications exacerbates the challenges of manual processing, resulting in delayed decision-making and inequitable distribution of financial aid.

Existing scholarship optimization tools despite their advantages have failed to address the existing challenges. Current predictive models such as the one proposed by Ahmad & Abu (2020) models have often struggled with accuracy on large, complex datasets. Furthermore, advanced techniques like deep learning remain underexplored despite their advantages when

dealing with large datasets with studies like a Fajardo et al. (2024) focusing on traditional models. Current research has been limited mainly to financial metrics, neglecting factors like student performance, financial need, bias, and fairness. Additionally, many studies lack model interpretability and focus narrowly on specific scholarship types, limiting broader applicability. There is a clear need for an automated, data-driven approach to scholarship award decisions that considers multiple variables and can scale with the increasing volume of applications. Deep learning algorithms, such as CNNs, provide a viable solution for analysing vast amounts of student data and predicting scholarship eligibility with high accuracy. This research developed a deep learning-based tool that streamlines the scholarship allocation process, improves decision accuracy, and ensures fairness in awarding financial aid.

1.3 Research Objectives

- i. To identify the challenges in the current scholarship allocation processes.
- ii. To review existing machine learning models and algorithms used for scholarship award decisions.
- iii. To develop a deep learning-based scholarship award decision tool using CNN.
- iv. To test and evaluate the developed tool.

1.4 Research Questions

- i. What are the primary challenges in the traditional scholarship allocation process?
- ii. What machine learning models and algorithms are currently being used in scholarship award decisions?
- iii. How can convolutional neural networks (CNNs) be used to develop an effective scholarship allocation tool?
- iv. How can the developed model be tested and validated?

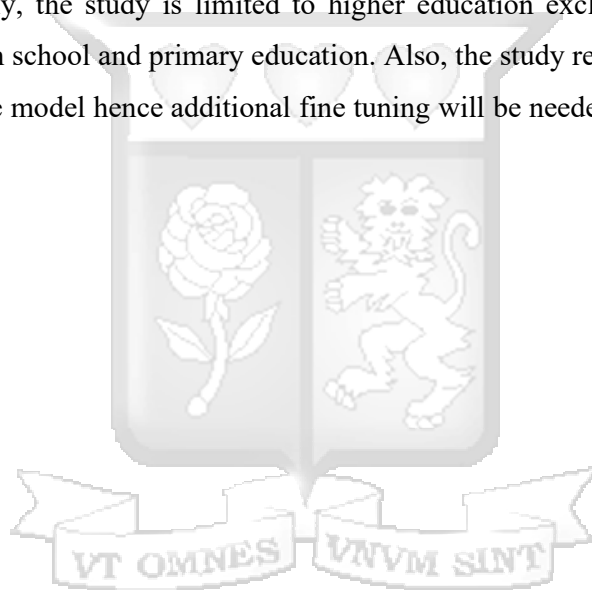
1.5 Justification

This research is significant as it addresses the inefficiencies and biases in the current scholarship allocation processes in higher education. By developing a deep learning tool that leverages CNNs, this study aims to provide a data-driven approach to scholarship award decisions, promoting fairness and transparency in the distribution of financial aid. This tool can help institutions make more informed decisions by considering a broader range of student data and reducing the reliance on manual, subjective evaluations (Chaudhari et al., 2022).

Furthermore, the findings of this research will contribute to the growing body of literature on AI applications in education and resource allocation, offering practical insights for policymakers and administrators.

1.6 Scope and Limitations

This study focuses on developing and testing a deep learning tool for scholarship allocation in higher education institutions. The scope is limited to the use of convolutional neural networks (CNNs) for predicting scholarship eligibility based on factors such as academic performance, financial need, and extracurricular involvement. The tool will be tested on historical student data from selected institutions to evaluate its accuracy and efficiency. The study however has few limitations. Firstly, the study is limited to higher education excluding other stages of education such as high school and primary education. Also, the study relies on secondary data to test and develop the model hence additional fine tuning will be needed to apply to different contexts.



Chapter 2: Literature Review

2.1 Introduction

In recent years, research on higher education scholarship award decisions has gained attention from both academia and industry. The selection of scholarship recipients has traditionally relied on a combination of academic performance, financial need, and other subjective factors, often leading to time-consuming and inconsistent results. With the advent of artificial intelligence (AI) and deep learning algorithms, these processes can now be automated and optimized for better accuracy and fairness (Li et al., 2022). The integration of deep learning models not only streamlines the decision-making process but also mitigates human bias, providing a more objective framework for evaluating candidates based on a wider array of factors. However, the adoption of AI tools in scholarship decisions raises critical questions around transparency, ethical considerations, and the potential exclusion of deserving candidates due to algorithmic biases (Zhao & Kim, 2023).

This chapter provides a comprehensive overview of the empirical and theoretical literature pertaining to scholarship award decisions in higher education. The chapter also explores the methodologies and findings from existing studies, theories underpinning the research, and challenges in current scholarship allocation processes. Additionally, the review examines the models and algorithms employed in these systems, alongside gaps in the literature and a conceptual model proposed for this research.

2.2 Challenges in the Current Scholarship Allocation Processes

The process of awarding scholarships in higher education institutions is fraught with several challenges, many of which have been highlighted in recent empirical and theoretical literature. These challenges stem from both the implementation of scholarship programs and the methodologies used to assess students' eligibility. The increasing reliance on data-driven models and machine learning algorithms has helped to mitigate some of these issues, but significant obstacles remain. Below are some of the key challenges that institutions face in the current scholarship allocation processes.

2.2.1 Inefficiency in Manual Processes

One of the primary challenges in scholarship allocation is the inefficiency of manual processes, which are often time-consuming and prone to human error. Many educational institutions still

rely on human evaluators to manually assess scholarship applications based on a range of criteria, including academic performance, financial need, and extracurricular activities (Orduña-Ferreira & Franzoni-Velazquez, 2024). The manual nature of this process can lead to inconsistencies in decision-making, as evaluators may apply subjective judgments or fail to consider all relevant factors due to time constraints. Additionally, manual processes are more susceptible to biases, such as favoritism or unintentional discrimination, which can affect the fairness of scholarship distribution (Ahmad & Bakar, 2020).

Automating the scholarship decision-making process through machine learning models and data-driven techniques has been shown to increase efficiency and reduce the likelihood of human error (Fajardo et al., 2024). However, many institutions lack the resources or expertise to implement such systems, particularly in lower-income regions or smaller universities. As a result, the inefficiencies of manual processes continue to be a major barrier to equitable and timely scholarship distribution (Dewi et al., 2020).

2.2.2 Bias and Inequity in Scholarship Allocation

Bias and inequity represent significant challenges in the scholarship allocation process. Despite efforts to create merit-based systems, research has shown that scholarships are often disproportionately awarded to students from more privileged backgrounds, who have better access to quality education and preparatory resources (Farias-Ruiz et al., 2023). This phenomenon is partly explained by the criteria used to evaluate applicants, which may inadvertently favor students from wealthier families or regions with better educational infrastructure (Littler, 2018). For example, standardized test scores and academic achievements, which are often key determinants in scholarship awards, can be influenced by socio-economic factors that are beyond the control of the students themselves.

To address these biases, some institutions have introduced need-based scholarships, which consider the financial background of students and attempt to level the playing field by providing additional support to those who are disadvantaged (Jordan et al., 2018). However, balancing merit and need remains a difficult challenge, as institutions must determine how to weigh different factors fairly while maintaining high standards of academic achievement (Jordan et al., 2018). Moreover, even need-based scholarships could suffer from biases if the data used to assess financial need is incomplete or inaccurate.

2.2.3 Lack of Standardization in Scholarship Criteria

One of the significant challenges in the scholarship allocation process is the lack of standardized criteria across institutions. Each higher education institution may have different benchmarks for awarding scholarships, such as varying academic thresholds, socio-economic considerations, or extracurricular achievements. This lack of standardization can make it difficult to compare and assess applicants fairly, particularly when students apply for multiple scholarships at different institutions (Jaisri et al., 2024). Additionally, inconsistencies in the criteria for merit-based and need-based scholarships can lead to confusion among applicants and evaluators alike. In some cases, highly deserving candidates may be overlooked due to the fragmented nature of the evaluation process. Standardizing the criteria for scholarship awards across institutions would allow for a more uniform and transparent process, reducing the subjectivity and bias inherent in the current system (Irfan et al., 2024). However, this would require significant coordination among educational institutions and governing bodies to establish common standards while still allowing for flexibility based on institutional priorities and student needs.

2.2.4 Difficulty in Measuring Holistic Student Potential

Another major challenge in scholarship award decisions is the difficulty in assessing students' holistic potential. While academic performance is a primary criterion, many scholarships also consider factors such as leadership skills, community service, and personal development. However, measuring these qualities objectively is challenging, as they are often subjective and difficult to quantify (Cosentino et al., 2019). Traditional evaluation methods, such as essays or recommendation letters, provide qualitative insights into a student's character and potential but are prone to bias and inconsistencies in how they are interpreted. Machine learning models that rely on quantitative data, such as grades and test scores, often fail to capture the full picture of a student's abilities. This challenge calls for the development of more sophisticated models that can account for both quantitative and qualitative data while maintaining fairness and objectivity (Cosentino et al., 2019). This will help institutions to find a balance between academic merit and personal achievements to ensure that scholarships are awarded to students who will benefit most from the opportunity.

2.3 Theoretical Literature

2.3.1 Human Capital Theory

Human Capital Theory, originally introduced by Becker (1994), posits that individuals invest in education and training to increase their productive capacity, leading to higher income and economic growth. The theory views education as a form of capital, like physical or financial capital, that contributes to the economic development of both individuals and society (Becker, 1994). In the context of higher education scholarships, Human Capital Theory provides a framework for understanding why scholarships are allocated to students based on their potential to contribute economically. Educational institutions, acting as gatekeepers of this human capital, allocate scholarships to individuals who are likely to generate the highest returns, typically measured in terms of academic success and employability.

The theory suggests that providing scholarships is an investment in human capital that benefits both the individual and society at large. By helping students access higher education, scholarships enable them to acquire skills and knowledge that enhance their future productivity. This, in turn, leads to greater economic returns in the form of higher wages and greater contributions to the labor market (Becker, 1994). However, one limitation of Human Capital Theory is its narrow focus on economic outcomes, often overlooking the broader social and personal benefits of education, such as increased civic engagement and personal well-being.

In scholarship award decisions, Human Capital Theory justifies merit-based scholarships to maximize societal returns on educational investments. The emphasis on academic performance and potential future contributions aligns with the theory's focus on maximizing economic returns. However, critics argue that this approach may exacerbate inequality by privileging students from already advantaged backgrounds who have greater access to resources and opportunities (Blaug, 1987).

2.3.2 Meritocracy Theory

Meritocracy Theory suggests that resources, opportunities, and rewards should be distributed based on merit, typically measured by individual ability and effort (Littler, 2017). In the context of higher education, meritocracy underpins the allocation of scholarships, with the assumption that those who demonstrate the most academic potential or effort deserve to be rewarded. This theory plays a significant role in the decision-making process for scholarships, as institutions

aim to identify students who have earned their place based on merit, whether through high academic performance, leadership, or other achievements. The meritocratic approach to scholarship allocation is often viewed as a way to promote fairness and equality of opportunity, as it allows individuals from diverse backgrounds to compete on an equal footing, provided they meet the required standards. However, critics of Meritocracy Theory argue that it overlooks structural inequalities that affect students' ability to compete fairly. For example, students from lower-income backgrounds may face significant barriers, such as lack of access to high-quality schooling or extracurricular opportunities that prevent them from achieving the same levels of "merit" as their more privileged peers (Kim & Choi, 2017).

In practice, the meritocratic approach to scholarship allocation may reinforce existing inequalities by rewarding students who have already benefited from advantageous social conditions (Mountford-Zimdars & Harrison, 2016). Therefore, while meritocracy remains a central principle in scholarship decisions, it is important to critically examine how merit is defined and whether it truly reflects individual potential rather than pre-existing advantages (Littler, 2018).

2.3.3 Equity Theory

Equity Theory, introduced by Adams (1965), focuses on fairness and justice in the distribution of resources. In the context of scholarships, Equity Theory suggests that scholarships should be distributed not only based on merit but also with consideration for individuals' unique circumstances, such as socio-economic background, access to educational resources, and other challenges. Equity in education is about ensuring that all students can succeed, even if it means providing additional support to those who face greater obstacles. Equity Theory plays an important role in shaping need-based scholarships, which are designed to address disparities in access to education by providing financial support to students from disadvantaged backgrounds. This approach contrasts with merit-based scholarships, which focus solely on academic achievement without considering students' socio-economic conditions (Mountford Zimdars, 2017).

Equity Theory argues that true fairness in scholarship allocation must account for the different starting points of students, recognizing that some students may require more resources to reach the same level of success as others (Martin & Kotzee, 2013). While merit-based scholarships

are grounded in the idea of rewarding individual effort, equity-based scholarships acknowledge that structural inequalities often limit certain students' ability to achieve high academic performance. By incorporating principles of equity into scholarship decisions, educational institutions can help level the playing field and ensure that all students have a fair chance to succeed

2.4 Empirical Literature

Scholarship award decisions have seen increasing attention from scholars who apply machine learning algorithms and data-driven methods to optimize selection processes. Several empirical studies have been conducted to analyse and predict the best candidates for scholarship allocation in higher learning institutions. Ahmad & Abu (2020) proposed an ensemble knowledge model aimed at supporting organizations in making scholarship award decisions. The model was designed to generate a list of eligible candidates, thereby reducing human error and the time taken to manually select candidates. Two ensemble approaches were presented: ensemble models and rule-based knowledge ensembles. The ensemble learning techniques used included boosting, bagging, voting, and rule-based ensemble techniques, alongside five base learners' algorithms, namely J48, Support Vector Machine (SVM), Artificial Neural Network (ANN), Naïve Bayes (NB), and Random Tree (RT). The study utilized a dataset containing 87,000 scholarship applications to train and evaluate the model. The ensemble voting technique achieved the highest accuracy of 86.9%, surpassing other techniques. The study also explored the rules generated from the J48 and Apriori models and used them to develop an improved rule-based ensemble classification model for scholarship awards.

However, the study's limitations include the potential overfitting of models due to the small dataset and the lack of generalization to different types of scholarships or educational institutions. Furthermore, the study did not evaluate the impact of varying applicant demographics or specific criteria changes on the model's performance. This research overcame the shortcomings of Ahmad & Abu's study. The developed model accuracy improved from 87% to 92% after fine tuning showcasing superior performance. The current research also used a very large sample size to mitigate the issues of bias and overfitting a challenge in Ahmad & Abu's research. Lastly, the current study focused on the impact of demographic characteristics on the influence of scholarship awards.

Aulck et al. (2020b) used data from a large public university to optimize scholarship disbursement and increase student enrolment. Initially, they predicted the enrolment of admitted first-year students using machine learning classifiers, achieving an AUROC of 0.85. These results were then combined with genetic algorithms to optimize the distribution of scholarships. Their model was projected to increase enrolment yield by 15.8%, and upon deployment, the university experienced a 23.3% increase in enrolment yield, resulting in significant tuition revenue gains. The university subsequently committed to using this model in future enrolment cycles. Despite the success of the model, the study had limitations, such as focusing solely on financial metrics and enrolment yield, without addressing the long-term academic success or retention of scholarship recipients. Additionally, the study's application was limited to one institution and put question to the generalizability of the results across different types of universities and demographic profiles. The current study addressed the generalizability challenge through carefully curated feature selection which allows for the model used across different various institutions.

Irfan et al. (2024) conducted a study to explore the use of machine learning algorithms, specifically Naive Bayes and Random Forest, to automate and streamline the scholarship distribution process. Their evaluation focused on accuracy, precision, and recall, demonstrating that data-driven approaches can promote fairness and inclusivity in scholarship awards. The findings underscored the potential for machine learning to enhance the efficiency of decision-making in educational settings. However, the study had notable limitations, including its narrow focus on two machine learning algorithms, which may not account for other potentially more effective models. Additionally, while the study emphasized fairness, it did not fully explore how these algorithms might handle biases in the data or how the models would perform in highly diverse student populations. Moreover, the study lacked real-world validation, as its findings were based on simulated data and scenarios, limiting its applicability to actual scholarship programs. In contrast, the CNN used in the current study offers superior capacity for complex data patterns and has been validated on real, diverse datasets. The present work also emphasizes interpretability challenges and fairness implications, which Irfan et al. only partially addressed.

Jaisri et al. (2024) proposed a sports scholarship prediction system using the K-Nearest Neighbours (KNN) algorithm to forecast scholarship amounts based on a student's sports achievements. The model was trained on a dataset incorporating academic and athletic

performance, with the KNN algorithm used to identify patterns and make predictions based on the closest matches in the data. The system provided a data-driven approach for predicting scholarships, offering educational institutions a method for making more informed decisions regarding sports scholarships. Although the system is simple and efficient, it is limited by its narrow focus on sports scholarships alone. This restricts its applicability to a wider range of scholarship programs. The current research addresses this gap by developing a broader model that accommodates diverse categories of scholarship applicants. Additionally, the KNN algorithm, while effective for small datasets, may struggle with large-scale applications due to its memory-intensive nature and decreased efficiency when processing high-dimensional data. Compared to this study, the selection of CNN model solves the challenge of working with large and high dimensional datasets which is critical for large volume of scholarship data. Furthermore, the study did not explore the potential for integrating non-performance-based factors, such as financial need. This research fully addresses this limitation by incorporating financial status, academic background, and sports participation into the model.

Fajardo et al. (2024) developed a predictive model for scholarship selection using a variety of machine learning algorithms, including Naïve Bayes, Random Forest, Logistic Regression, Support Vector Machine (SVM), and Multilayer Perceptron. The study aimed to match applicants with the most suitable scholarships based on their profiles and qualifications. The researchers evaluated the model's using metrics like accuracy, precision, recall, and F1-score, with Logistic Regression emerging as the best-performing model in terms of overall accuracy and balance. SVM, Naive Bayes, and Random Forest also showed competitive results, while the Multilayer Perceptron performed the worst.

Despite these findings, the study faced several limitations. It did not explore the interpretability of the models, which could be critical for transparency in scholarship decisions. Additionally, the study's focus on algorithm performance overlooked other factors that might influence scholarship distribution, such as ethical concerns around bias and fairness. The research also did not assess the models' performance across varying types of scholarships or different institutional requirements, limiting its broader applicability. The present research addresses these gaps by not only achieving high accuracy with a deep learning model but also conducting detailed error analysis through confusion matrices and classification reports to uncover potential biases and limitations.

Table 2.1 Summary of the Empirical Studies

Author	Techniques	Accuracy	Limitations
Ahmad & Abu (2020)	Ensemble Learning (Boosting, Bagging, Voting, Rule-based ensembles), J48, SVM, ANN, Naive Bayes, Random Tree	86.9%	<ul style="list-style-type: none"> -Limited evaluation of varying demographics or scholarship types. -The research used small dataset which could affected the accuracy of the models.
Aulck et al. (2020b)	Machine learning classifiers combined with Genetic Algorithms	AUROC of 0.85	<ul style="list-style-type: none"> -Focused on financial metrics and enrolment yield leaving other metrics such as performance -The model has limited generalizability across institutions and demographics.
Irfan et al. (2024)	Naive Bayes, Random Forest	Not specified	<ul style="list-style-type: none"> -Narrow focus on two algorithms excluding advanced algorithms such as deep learning. -The model lacks bias handling. -The research utilized simulated data which may have applicability issues.
Jaisri et al. (2024)	K-Nearest Neighbours (KNN)	Not specified	<ul style="list-style-type: none"> - The research focused only on sports scholarships -The algorithms used are memory intensive. -The research ignored non performance metrics such as financial need -The dataset for this research was relatively small.

Fajardo et al. (2024)	Naive Bayes, Random Forest, Logistic Regression, SVM, Multilayer Perceptron	Not specified (Logistic Regression performed best overall)	-The research did not explore model interpretability. -The study overlooked bias and fairness -The research focused only on algorithm performance.
-----------------------	---	--	--

2.5 Existing Machine Learning Models and Algorithms for Scholarship Award

Decisions.

As the demand for efficient and fair scholarship allocation grows, numerous models and frameworks have been employed to automate and optimize the decision-making process. These models rely on machine learning and data mining techniques to predict and evaluate scholarship eligibility based on academic performance, financial need, and other relevant factors. In this section, we explore some of the key models and frameworks used in scholarship allocation systems, along with their strengths and limitations.

2.5.1 Models

Several machine learning models have been implemented in scholarship award decisions to improve accuracy, efficiency, and fairness. The following subsections provide an overview of the most used models.

2.5.1.1 Ensemble Model

Ensemble learning involves training multiple models and combining their outputs, treating them like a "committee" of decision-makers (Langford et al., 2011). The idea is that the collective decision, when the individual predictions are combined appropriately, typically results in better accuracy than any single model on its own. Numerous empirical and theoretical studies show that ensemble models often outperform individual models in terms of accuracy. The models in an ensemble can predict various types of outputs, such as continuous values, class labels, probabilities, rankings, or clusters. Their decisions can be combined through techniques like averaging, voting, or probabilistic methods. Most ensemble learning techniques are versatile and can be applied to a wide range of model types and learning tasks.

The ensemble model is a powerful machine learning approach that combines multiple base learners to improve predictive performance. Ensemble techniques such as bagging, boosting, and stacking allow for better generalization and robustness compared to individual models (Mohammed & Kora, 2023). Ensemble models work by aggregating the predictions of several models to produce a final prediction, which reduces the likelihood of overfitting and improves accuracy. In scholarship allocation, ensemble models have been employed to integrate various criteria, including academic performance, financial need, and extracurricular achievements.

Ahmad and Abu Bakar (2020) applied ensemble learning techniques, including bagging and boosting, to predict scholarship eligibility with high accuracy. Their study found that ensemble models significantly outperformed single classifiers such as Support Vector Machines (SVM) and Naïve Bayes, with accuracy rates reaching 86.9%. The main advantage of the ensemble model is its ability to combine the strengths of different algorithms, making it particularly effective for complex, multi-criteria decisions. Despite their high accuracy, ensemble models require significant computational resources, making them less suitable for institutions with limited technical infrastructure. Additionally, training and maintaining ensemble models can be time-consuming, particularly when dealing with large datasets (Fajardo et al., 2024).

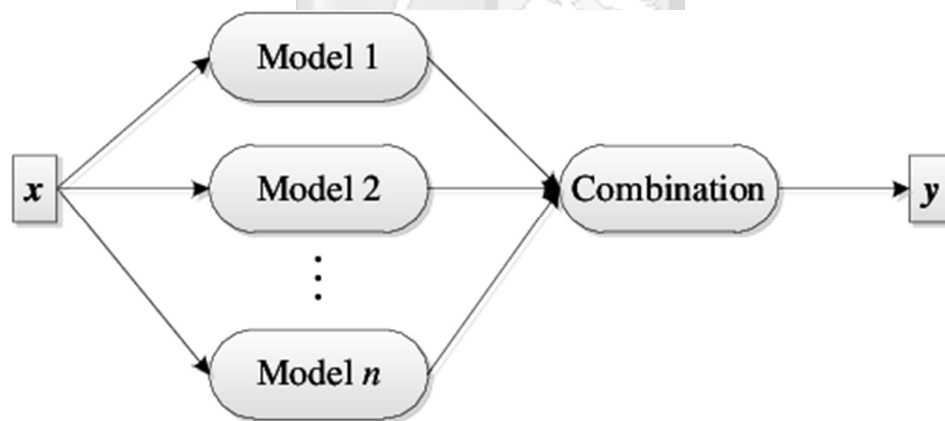


Figure 2.1 Architecture of Ensemble Model (Petrankova et al., 2015).

2.5.1.2 Random Forest Model

Random forests are an ensemble learning method made up of multiple decision trees, where each tree is built using a random selection of features from the dataset (Breiman, 2001). Each tree in the forest is trained on a random vector that is sampled independently but follows the same distribution for all trees. As the number of trees in the forest increases, the generalization

error of the forest converges to a stable limit. The overall performance of a random forest depends on two factors: the strength (or accuracy) of individual trees and the correlation between them. Lower correlation among trees and stronger individual trees contribute to better performance. By randomly selecting features to split nodes, random forests tend to have error rates comparable to those of Adaboost, as noted by Breiman (2001) but with the added advantage of being more resilient to noisy data. Random forests also provide internal estimates to monitor metrics such as error rate, individual tree strength, and the correlation between trees. These estimates help assess the effect of using more features for splitting and can be used to measure the importance of variables in the model. This approach applies not only to classification tasks but also to regression problems, where random forests can provide robust predictions.

Gani et al. (2024) used Random Forest models to predict scholarship eligibility based on academic performance, financial status, and extracurricular activities. The Random Forest model achieved an accuracy of 93%, outperforming other machine learning models such as Naïve Bayes and Logistic Regression. The ability of Random Forest to handle complex datasets and large numbers of features made it well-suited for the scholarship allocation process, where multiple criteria need to be considered simultaneously. Although Random Forest is highly accurate and robust, it can be computationally expensive, particularly when applied to large datasets. Moreover, the interpretability of the model is relatively low compared to simpler models like decision trees, making it difficult for decision-makers to understand how predictions are made (Ahmad & Bakar, 2020).

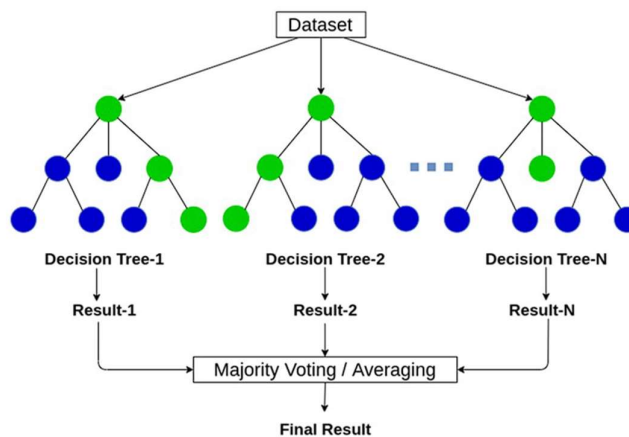


Figure 2.2 Random Forest Model Architecture (Mumtaz et al., 2023).

2.5.1.3 Support Vector Model (SVM)

Support Vector Machines (SVM) are a robust technique for building classifiers. The goal is to establish a decision boundary between two classes to predict labels from one or more feature vectors (Huang, 2018). This boundary, called the hyperplane, is positioned in such a way that it maximizes the distance from the nearest data points of each class. These closest data points are referred to as support vectors.

Consider a labelled training dataset:

$(x_1, y_1), \dots, (x_n, y_n)$, where $x_i \in \mathbb{R}^d$ represents a feature vector, and $y_i \in \{-1, +1\}$ is the class label (negative or positive) for training sample i .

The optimal hyperplane can be expressed as:

$$w^T x + b = 0$$

where w is the weight vector, x is the input feature vector, and b is the bias term.

The weight vector w and bias b must satisfy the following conditions for all elements in the training set:

$$w^T x_i + b \geq +1 \text{ if } y_i = 1$$

$$w^T x_i + b \leq -1 \text{ if } y_i = -1$$

The purpose of training an SVM model is to determine w and b such that the hyperplane both separates the data and maximizes the margin, which is calculated as $1 / \|w\|^2$. Data points x_i for which $|y_i| (w^T x_i + b) = 1$ are known as support vectors. The Support Vector Model (SVM) is a supervised learning algorithm commonly used for classification tasks. It works by finding the optimal hyperplane that best separates different classes in the dataset. In the context of scholarship allocation, SVM is used to classify students into eligible or non-eligible categories based on various features such as academic performance, financial need, and socio-economic background.

Muhammad et al. (2023) used SVM in their scholarship allocation system and found it to be effective for smaller datasets, with an accuracy of 84%. SVM's strength lies in its ability to handle high-dimensional data and its robustness to overfitting, especially in cases where the number of features exceeds the number of data points. While SVM performs well with smaller datasets, it can struggle with scalability, particularly when applied to large datasets with thousands of students. Additionally, the algorithm's performance depends heavily on the choice of kernel and selecting the appropriate kernel function can be challenging (Ahmad & Bakar, 2020).

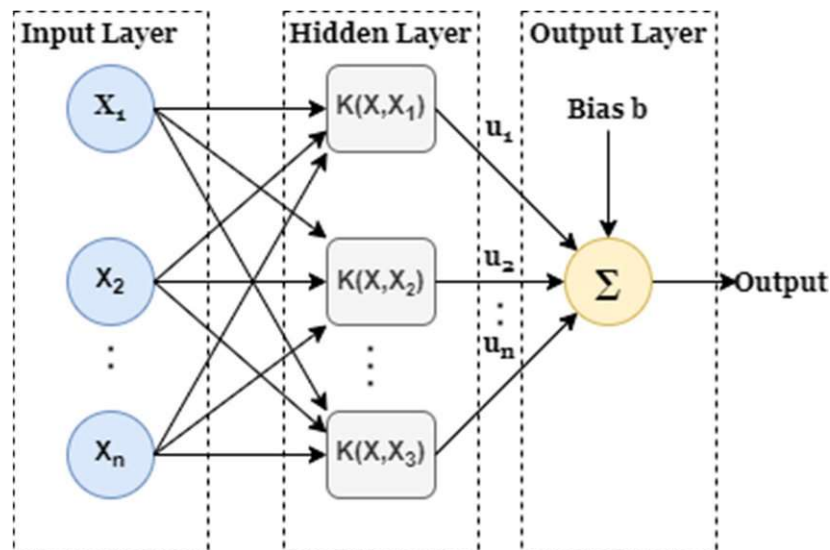


Figure 2.3 SVM Model Architecture (Kadam et al., 2021).

2.5.1.4 Multilayer Perceptron

The multilayer perceptron (MLP) is a type of feed-forward neural network consisting of three main layers: an input layer, a hidden layer, and an output layer, as illustrated in Figure 2.4. The input layer receives the data to be processed, while the output layer handles tasks such as classification and prediction (Chai et al., 2021). The MLP features a series of hidden layers between the input and output layers, through which data flows in a forward direction, similar to a feed-forward network. Training of the MLP is done using the backpropagation algorithm, which updates the weights of the nodes. MLPs are capable of solving problems that are not linearly separable and are designed to approximate any continuous function.

MLP is particularly well-suited for complex decision-making tasks where multiple inputs and non-linear relationships exist between variables. In scholarship allocation systems, MLP can

model the interactions between academic, financial, and socio-economic factors to determine eligibility for scholarships. Dewi et al. (2020) applied MLP to their scholarship allocation system, achieving a high level of accuracy by modelling the non-linear relationships between input features. MLP's ability to learn complex patterns in the data made it effective for predicting scholarship eligibility, particularly when multiple criteria were considered simultaneously. One of the main challenges with MLP is the need for a large amount of data to train the model effectively. Additionally, MLP models are often difficult to interpret, which can be a disadvantage when transparency is required in the scholarship decision-making process (Ahmad & Bakar, 2020).

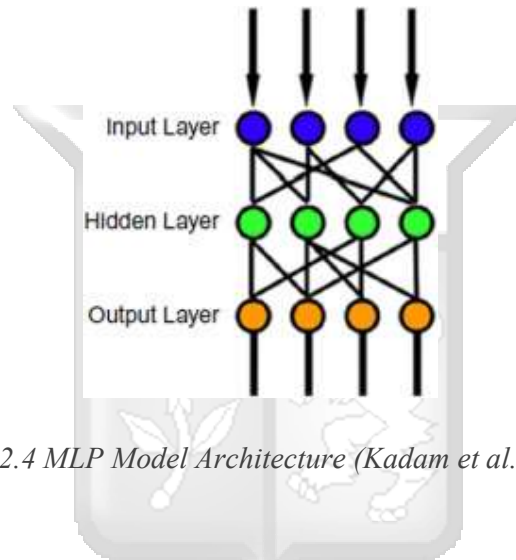


Figure 2.4 MLP Model Architecture (Kadam et al., 2021).

2.5.2 Frameworks

2.5.2.1 TensorFlow

TensorFlow is an open-source machine learning framework developed by Google. The initial version of TensorFlow, released in 2015, aimed to integrate various libraries into a unified package capable of solving a wide range of machine learning problems (Novac et al., 2022). This approach provided significant flexibility in network model design, with libraries like Contrib and Keras available to define layers. While the framework offered all the necessary tools for experienced users, the inclusion of multiple libraries sometimes led to more complex interactions and made certain aspects harder to understand. For example, users needed to write separate codes for training the network, making it difficult to identify and correct mistakes. As a result, TensorFlow was initially perceived as less user-friendly. However, the release of TensorFlow 2.0 in 2019 addressed many of these issues (Novac et al., 2022).

This updated version introduced major design changes, including the removal of certain libraries to reduce system complexity and confusion. To improve the training process and align

with competing platforms, the previous two-step training process was simplified, with the second step now handled internally, making it easier to trace. Additionally, the new version streamlined interaction by introducing a more intuitive instantiation process, utilizing Python-based user interfaces like classes to define core components. In addition to resolving earlier issues, TensorFlow 2.0 introduced new features such as the TensorFlow Datasets module, which provided a simpler way to feed data into the network, and a new runtime abstraction that allowed networks to be trained on local or multi-GPU environments without any modifications.

These improvements made TensorFlow 2.0 popular among both experienced and novice users, offering strong performance along with an easier-to-use interface that significantly reduced the learning curve. It is widely used for building deep learning models, including those used in scholarship allocation systems. TensorFlow's flexibility and scalability make it ideal for handling large datasets and complex models such as MLP and CNN (Convolutional Neural Networks). TensorFlow also provides tools for distributed computing, which allows institutions to process large volumes of scholarship applications efficiently.

2.5.2.2 Scikit-Learn

Scikit-Learn is a Python-based machine learning library that provides simple and efficient tools for data mining and analysis (Pedregosa et al., 2011). It is particularly useful for implementing basic machine learning algorithms such as decision trees, random forests, and SVM. Scikit-Learn's ease of use and versatility make it a popular choice for educational institutions seeking to implement machine learning models without the need for advanced deep learning techniques.

2.5.2.2 Keras

Keras is an open-source neural network library that provides a user-friendly interface for building and training deep learning models. Keras is designed to be easy to use, allowing developers to quickly build models without needing to write complex code. It is often used in conjunction with TensorFlow for more advanced machine learning tasks.

2.6 Architectures and Designs

The design and architecture of scholarship allocation systems are critical in determining their scalability, accuracy, and ease of use. Various architectures, from data mining systems to advanced artificial intelligence frameworks, have been proposed and tested in academic

research to optimize the scholarship award process. This section discusses some of the most common architectures used in scholarship prediction models, including data mining architectures, Naïve Bayes and Random Forest architectures, holistic designs, and specific frameworks tailored for educational institutions.

2.6.1 Data Mining Architecture

Data mining architecture has been widely adopted in scholarship allocation systems for their ability to sift through large datasets and identify patterns and relationships that can predict scholarship eligibility. These architectures typically involve three stages: data pre-processing, pattern discovery, and result evaluation. The data pre-processing stage cleanses and organizes the data, ensuring that it is in a format suitable for mining. In the pattern discovery phase, machine learning algorithms such as decision trees, clustering, and association rules are applied to uncover hidden patterns. Finally, the results are evaluated to determine the accuracy and relevance of the predictions. Figure 2.5 from Delima (2019) illustrates data mining architectural framework of the study. The data collected from the Student Financial Assistance Unit office was reprocessed, and selected attributes were used for clustering. The clustered scholarship grants were then used to forecast the number of future grantees.

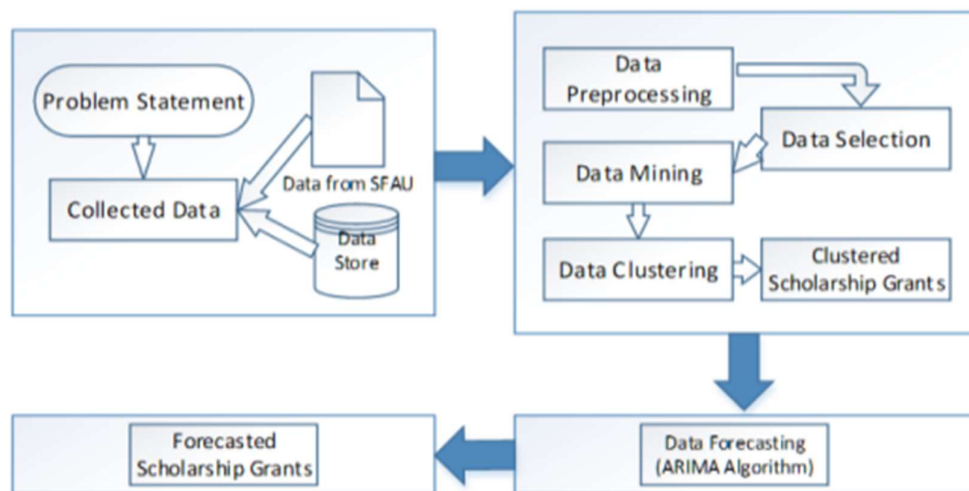


Figure 2.5 Data Mining Architecture (Zhou et al, 2020)

2.6.2 Naive Bayes and Random Forest Architecture

The combination of Naïve Bayes and Random Forest architectures has gained popularity in scholarship prediction systems due to the complementary strengths of these models. Naïve

Bayes is a probabilistic model that is well-suited for handling large datasets with many features, while Random Forest is an ensemble model that builds multiple decision trees to improve prediction accuracy. The integration of these two models allows the system to take advantage of Naïve Bayes' simplicity and Random Forest's robustness, resulting in a more efficient and accurate scholarship allocation process.

The architecture proposed by Irfan et al. (2024) utilizes machine learning algorithms, such as Naive Bayes and Random Forest, to efficiently match scholarships with deserving students based on predefined criteria. By integrating data from both scholarship and student profiles, the system enables highly accurate and fair identification of suitable matches. This architecture provides scholarship providers and educational institutions with a streamlined and objective method for scholarship allocation, fostering inclusivity and equity in the distribution of educational opportunities.

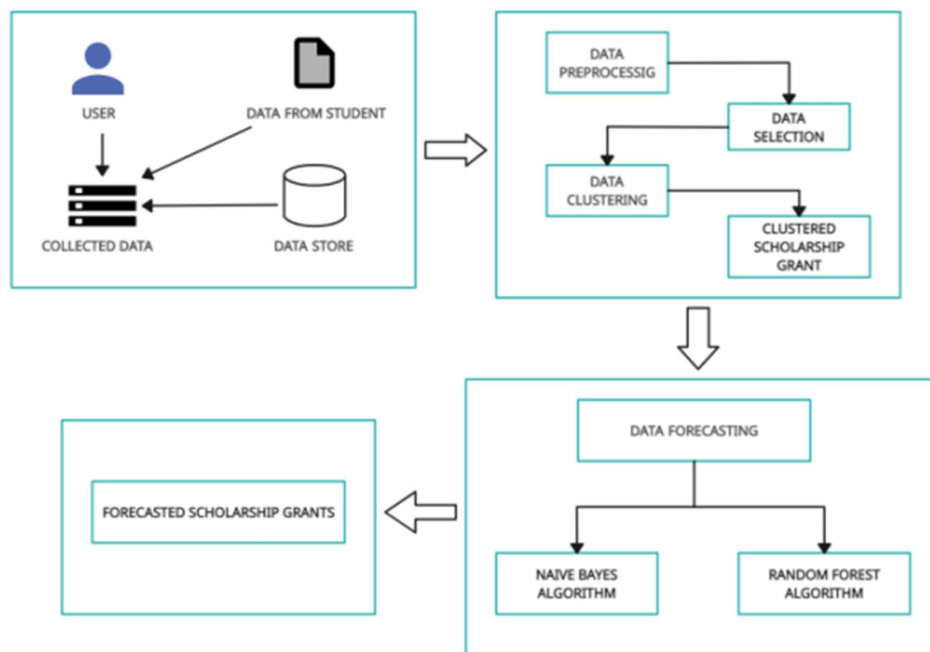


Figure 2.6 Naive Bayes and Random Forest Architecture (Irfan et al., 2024).

2.6.3 Holistic Architecture

Holistic architecture in scholarship allocation systems aim to capture a comprehensive view of student potential by integrating both quantitative and qualitative data. In addition to academic performance and financial need, holistic architecture may consider factors such as leadership

skills, community involvement, and personal development. This approach is particularly useful in institutions that prioritize a well-rounded view of students rather than focus solely on academic merit. Jaisri et al. (2024) presented holistic architecture that outlines the training phase, explaining how the model learns from the dataset, recognizes patterns, and establishes relationships among various features. The testing phase is then explored, detailing the process of uploading a sports certificate, extracting relevant information, and applying the trained KNN model for prediction. The architecture emphasizes the transparency and interpretability of the KNN algorithm throughout the process.

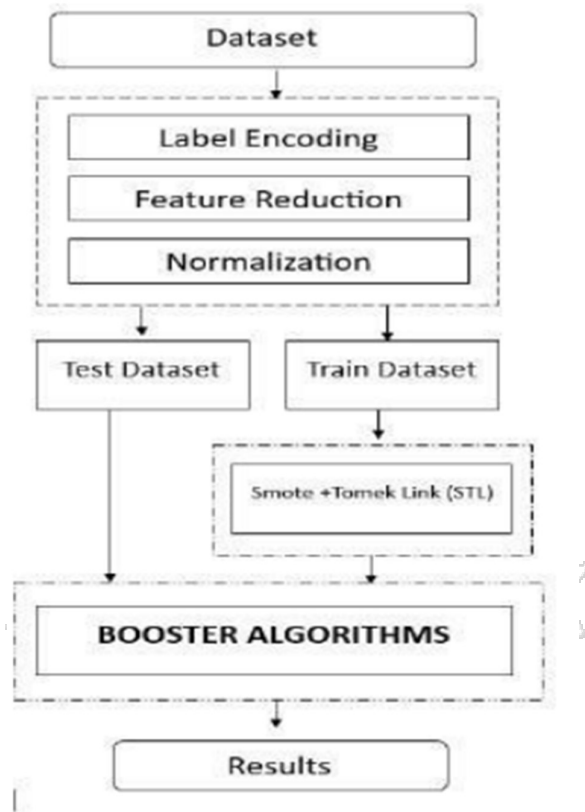


Figure 2.7 Holistic Architecture (Irfan et al., 2024).

2.6.4 Student Scholarship prediction Architecture

The Student Scholarship Prediction Architecture is specifically designed to streamline the scholarship allocation process in educational institutions. This architecture integrates data from multiple sources, including academic records, financial aid applications, and extracurricular activities, to build a comprehensive profile of each student. Machine learning algorithms are then applied to predict scholarship eligibility based on predefined criteria. The architecture proposed by Chaudhari et al. (2022) demonstrated the input provided by users and the

corresponding output generated by the system. In the student module, various attributes related to students were considered to assess their performance and determine scholarship eligibility. Key attributes included numerical parameters such as 10th and 12th grade marks, and marks from the 1st, 2nd, and 3rd years of college.

The raw data used for this study was collected from a reputable engineering college and included attributes such as 10th and 12th grade marks, 1st, 2nd, and 3rd year marks, diploma results, gaps, and ATKT (Allowed to Keep Terms). The dataset consisted of 150-200 records, with 70% allocated for training and 30% for testing. Data Preprocessing and Feature Extraction: The data was preprocessed to ensure completeness, and the selected fields for the model included 10th and 12th marks, 1st, 2nd, and 3rd year marks, diploma marks, gaps, ATKT, and result. These variables were based on prior student performance data, ensuring the reliability and validity of the model. Classification: The study utilized machine learning and classification algorithms. A two-layered classifier system was designed, chosen for its effectiveness in various domains. Naive Bayes was used due to its strong predictive accuracy, learning capability, and ability to retain information, making it well-suited for real-world classification tasks across industries, business, science, and education.

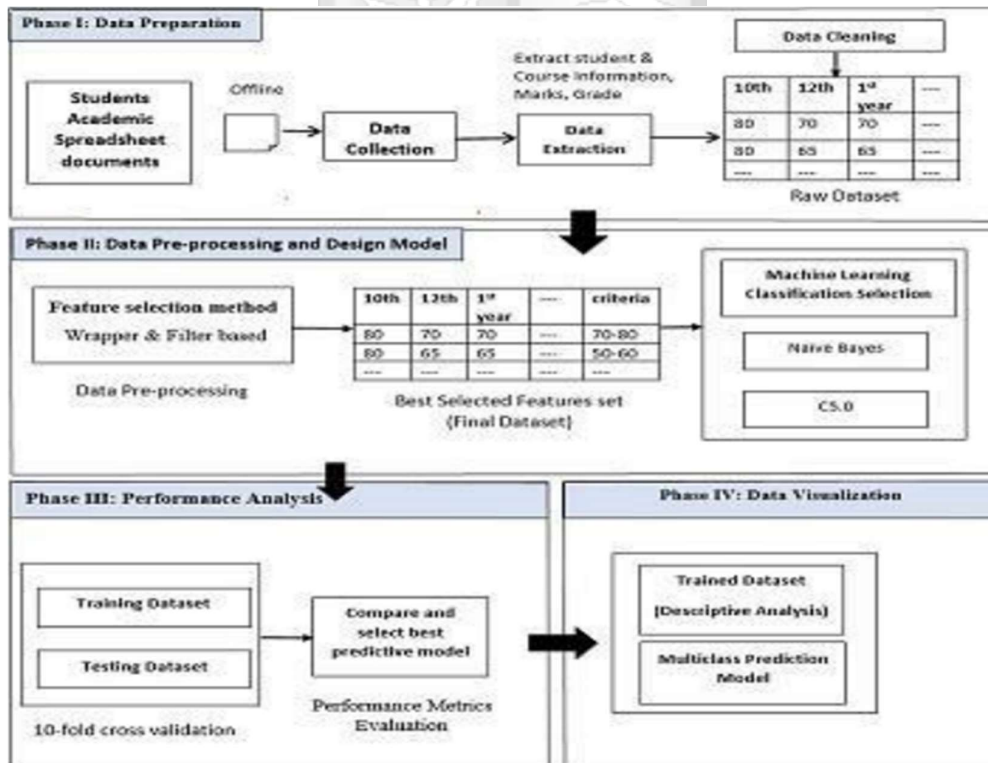


Figure 2.8 Holistic Architecture (Irfan et al., 2024)

2.7 Algorithms

The success of scholarship allocation systems largely depends on the algorithms employed to predict student eligibility and distribute resources equitably. These algorithms vary in complexity and efficiency, depending on the size and quality of the dataset, the criteria used for scholarship decisions, and the computational resources available to the institution. In this section, we will discuss key algorithms used in scholarship allocation systems, their mechanisms, strengths, and limitations.

2.7.1 ARIMA algorithm

ARIMA (AutoRegressive Integrated Moving Average) is a statistical analysis technique that is often used for time series forecasting. While typically applied to financial and economic data, ARIMA can be adapted for educational contexts, particularly in tracking student performance trends over time to inform scholarship allocation decisions. ARIMA models use historical data to predict future trends, making them useful for scholarship systems that require ongoing performance evaluations of students throughout their academic careers (Wei et al., 2016). ARIMA models require continuous and high-quality time series data, which may not be available in many educational institutions. Additionally, ARIMA's forecasting capabilities are limited to linear relationships, making it less suitable for complex, multi-criteria scholarship decisions.

2.7.2 Genetic Algorithms

Genetic algorithms (GAs) are search methods inspired by the principles of natural selection and genetics (Fraser, 1957; Bremermann, 1958). GAs encode the decision variables of a search problem into finite-length strings of symbols, referred to as chromosomes. These chromosomes represent candidate solutions, with each symbol acting as a gene, and the values of the genes called alleles. For instance, in the traveling salesman problem, a chromosome represents a route, and a gene represents a city. Unlike traditional optimization techniques, GAs work with coded parameters rather than the parameters themselves.

To evolve high-quality solutions through natural selection, a measure to differentiate good solutions from bad is necessary. This measure can be an objective function modeled mathematically or through a computer simulation, or it could be a subjective function where humans choose better solutions over worse ones. Fitness measure determines a solution's

relative fitness and guides the GA in evolving optimal solutions. A key concept in GAs is the population. Unlike traditional search methods, GAs operates on a population of candidate solutions. The population size, usually defined by the user, greatly impacts the scalability and performance of the algorithm. A small population may lead to premature convergence and suboptimal solutions, while a large population can increase computational demands. Once the problem is encoded in chromosomes and a fitness measure is chosen, the GA evolves solutions through the following steps:

- i). **Initialization:** The initial population is typically generated randomly across the search space, though domain-specific knowledge can be integrated.
- ii). **Evaluation:** Once initialized, or after an offspring population is created, the fitness of each candidate solution is evaluated.
- iii). **Selection:** Selection assigns more copies to candidate solutions with higher fitness values, following the "survival of the fittest" principle. Several selection methods exist, such as roulette-wheel selection, stochastic universal selection, ranking selection, and tournament selection.
- iv). **Recombination:** Recombination mixes parts of two or more parental solutions to generate new, potentially better offspring. Performance depends on an effective recombination mechanism, and offspring typically combine traits from parents in novel ways (Goldberg, 2002).
- v). **Mutation:** Mutation randomly modifies an individual solution. While recombination involves multiple parents, mutation introduces local random changes to one or more traits of a candidate solution.
- vi). **Replacement:** The new population, created by selection, recombination, and mutation, replaces the original parent population. Techniques like elitist replacement, generation-wise replacement, and steady-state replacement are commonly used.
- vii). **Repeat:** Steps 2-6 are repeated until a specified termination condition is met.

This iterative process continues until the GA evolves optimal or near-optimal solutions to the search problem.

2.7.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that is widely used for classification tasks, including scholarship allocation. SVM works by finding the optimal

hyperplane that separates different classes in the dataset (e.g., eligible and non-eligible students). SVM is particularly effective in cases where there are clear boundaries between classes, making it useful for distinguishing between high-performing students and those who may not meet scholarship criteria.

2.7.4 Artificial Neuron Network (ANN)

Artificial Neural Networks (ANNs) are modeled after the functioning of biological neural systems, such as how the brain processes information (Grosan & Abraham, 2011). These systems consist of numerous interconnected processing units, called neurons, which work together to solve specific problems. Like humans, ANNs learn through examples. Like learning in biological systems, ANN learning involves adjusting the synaptic connections between neurons. The concept of the first artificial neuron was introduced in 1943 by neurophysiologist Warren McCulloch and logician Walter Pitts, although the technology at that time was insufficient to advance their work significantly. There is a notable distinction between how neural networks solve problems compared to traditional algorithms.

In conventional algorithms, the computer follows a defined set of instructions to solve a problem. These steps are explicit, and without them, the computer cannot function. However, computers could be even more valuable if they could perform tasks that humans do not fully understand how to execute. In traditional computing, the solution to a problem must be clearly defined in a series of precise instructions, which are then translated into a high-level programming language and subsequently into machine code for the computer to process. These systems are highly predictable, and any issues typically stem from software or hardware failures (Grosan & Abraham, 2011).

Neural networks, on the other hand, cannot be programmed to perform specific tasks directly; instead, they learn by example. It is crucial to select examples carefully, as poor selection can lead to wasted time or improper network functioning. Since ANNs determine how to solve problems independently, their behavior can sometimes be unpredictable (Katoch et al., 2020). ANNs and conventional algorithms are not in competition but are complementary. Certain tasks are more suited to an algorithmic approach, while others are better suited to neural networks. Many tasks require a combination of both approaches for optimal efficiency. For instance, a conventional computer might be used to supervise a neural network to achieve maximum performance (Grosan & Abraham, 2011). ANNs are highly effective at modelling

complex relationships between variables and can be used to predict scholarship eligibility based on multiple inputs, such as grades, financial data, and personal achievements. The flexibility of ANNs allows them to learn from both structured and unstructured data, making them suitable for holistic scholarship systems.

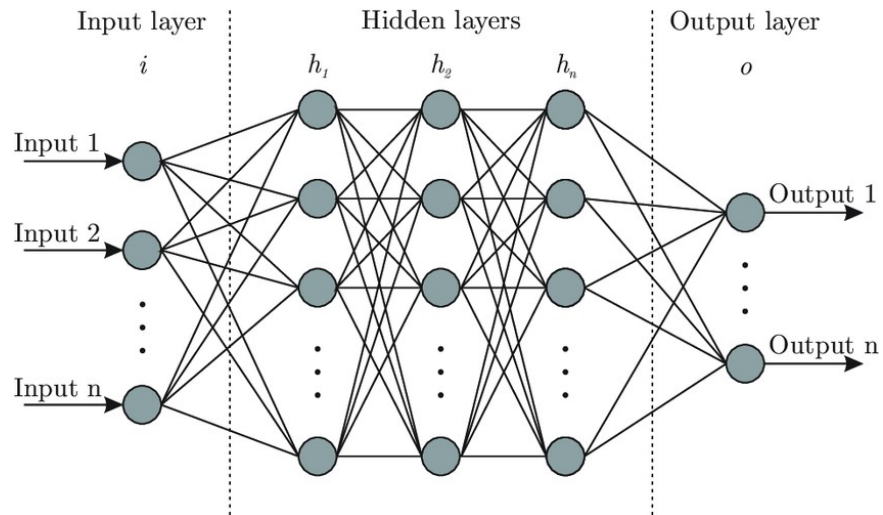


Figure 2.9 Structure of ANN (Gurucharan, 2020).

2.7.5 Naïve Bayes

The Naive Bayes Classifier is a type of probabilistic classifier based on Bayes' theorem. It operates on the assumption that the presence of a particular feature in a class is independent of the presence of other features in the same class (Zhang & Gao, 2011). For instance, when classifying a fruit as a watermelon, the classifier considers characteristics like being green, round, and having a diameter of about 15 cm as independent features. In this case, a Naive Bayes model would calculate the probability of the fruit being a watermelon based on each of these features separately. For example, there might be a 60% probability that a fruit is a watermelon if it is green, a 40% probability if it is round, and a 95% probability if its diameter is around 15 cm. Since the model treats each feature independently in the probability calculation, it is referred to as "Naive."

According to Bayes' theorem:

$$p(C_j | x) = \frac{p(C_j)p(x|C_j)}{p(x)} \text{ Equation 2.1 Bayes Theorem}$$

Where:

C_j Represents the number of classes, and x is the feature vector.

$p(C_j | x)$ is the posterior probability of class C_j given the predictor x .

$p(C_j)$ is the prior probability of the class.

$p(x|C_j)$ is the likelihood, or the probability of the predictor given a class.

$p(x)$ is the prior probability of the predictor.

Naïve Bayes is often used in scholarship allocation systems to classify students based on academic performance, financial need, and other independent variables.

2.7.6 K-Nearest Neighbors (KNN)

K-nearest-neighbour (K-NN) is one of the most fundamental and effective algorithms for data classification, particularly useful when the data is ambiguous. The algorithm was first introduced in 1951 by Evelyn Fix and Joseph Hodges for discriminant analysis, addressing the challenges of estimating probabilistic densities through parametric methods (Fix, 1985). In 1967, additional properties of the algorithm were studied, such as when 'k' equals 1 and 'n' approaches infinity, the classification error of K-NN is bounded above by twice the Bayes error rate. Following the discovery of these characteristics, extensive research was conducted over the years to explore new rejection methods, improve Bayes error rates, develop distance-based techniques, soft computing methods, and other enhancements (Bansal et al., 2021).

K-NN is classified as a supervised learning algorithm and is considered one of the simplest and most user-friendly algorithms in machine learning. Although it can be used for both classification and regression, it is primarily applied for classification tasks. K-NN is highly useful for imputing missing values and resampling data (Bansal et al., 2022). For a given dataset, the algorithm predicts the relationship between new data and the existing data, assigning the new data point to the category that best matches its neighbours. Thus, the K-NN algorithm effectively classifies fresh data by evaluating its proximity to neighbouring points. K-NN is often referred to as a "lazy learner" because it does not perform any training on the dataset until a classification or prediction is needed. The algorithm is also non-parametric, meaning it does not assume any predefined relationship between input and output variables. In scholarship allocation systems, KNN works by comparing a student's features (e.g., grades,

financial status) to the features of other students in the dataset. The algorithm assigns the student to the class (eligible or non-eligible) based on the majority class of its nearest neighbors.

Table 2.2 shows the summary of the algorithms that have been used in Higher Learning Scholarship Award Decisions.

Table 2.2 Summary of Algorithms.

Algorithm	Strengths	Limitations
ARIMA	Useful for time series forecasting	Limited to linear relationships, requires continuous data
Genetic Algorithm	Optimal for multi-objective optimization	Computationally expensive, may converge on local optima
Support Vector Machine (SVM)	Effective for high-dimensional data	Struggles with scalability, kernel optimization issues
Artificial Neural Networks	Models' complex relationships, handles large data	Requires large datasets, difficult to interpret
Naïve Bayes	Fast and efficient	Assumes feature independence, can lead to inaccuracies
Random Forest	High accuracy, handles large datasets	Computationally expensive, low interpretability
K-Nearest Neighbours (KNN)	Simple and easy to implement	Computationally expensive, affected by noisy data
Logistic Regression	Simple and interpretable	Assumes linear relationships, struggles with non-linear data

2.8 Gaps in Literature

Despite significant advancements in scholarship allocation systems, a critical research gap exists in developing accurate, scalable, and fair predictive models. Current approaches, such as Random Forest and Support Vector Machines (SVM), perform well on small datasets but struggle to maintain accuracy and scalability when applied to larger, more diverse datasets

(Ahmad & Bakar, 2020). Additionally, existing models predominantly focus on financial metrics and enrolment yield, neglecting essential factors such as student performance, financial need, and socio-economic disparities that are critical for equitable scholarship allocation (Fajardo et al., 2024). Most studies have relied on traditional machine learning techniques, overlooking the potential of advanced methods like deep learning, which can handle complex and large-scale data more effectively. Moreover, the lack of model interpretability limits stakeholder trust, and issues of bias and fairness remain inadequately addressed.

To bridge this gap, this research aims to develop a CNN-based algorithm that evaluates scholarships using a comprehensive set of input parameters, including financial need, academic performance, and personal background. A key feature of the proposed model will be its ability to optimize scholarship scores in cases where candidates tie, ensuring fair and transparent allocation decisions. By integrating interpretability and addressing socio-economic disparities, the research will contribute to the development of an innovative, scalable, and equitable scholarship allocation system, setting it apart from existing methodologies.

2.9 Conceptual Model

The conceptual model for the proposed tool is illustrated in Figure 2.10, showcasing the key stages involved in model training. The first step in the training phase is data preparation, where scholarship data is preprocessed to make it suitable for the machine learning model. This step involves feature extraction, label encoding, and normalization to ensure the data is structured and scaled properly. Once the data is prepared, it is split into two subsets: a training set and a test set. During the model training phase, the Convolutional Neural Network (CNN) algorithm is trained using the training set. The CNN learns the patterns and relationships within the data to create a predictive model. After the training process is complete, the model is evaluated using the test set to measure its performance and accuracy. Once the model has undergone testing, it proceeds to the validation phase. The validation process ensures the model's predictions are reliable and robust. After successful validation, the model is ready for deployment and can be used to predict scholarship eligibility for applicants, effectively providing insights based on the input features from the prepared data.

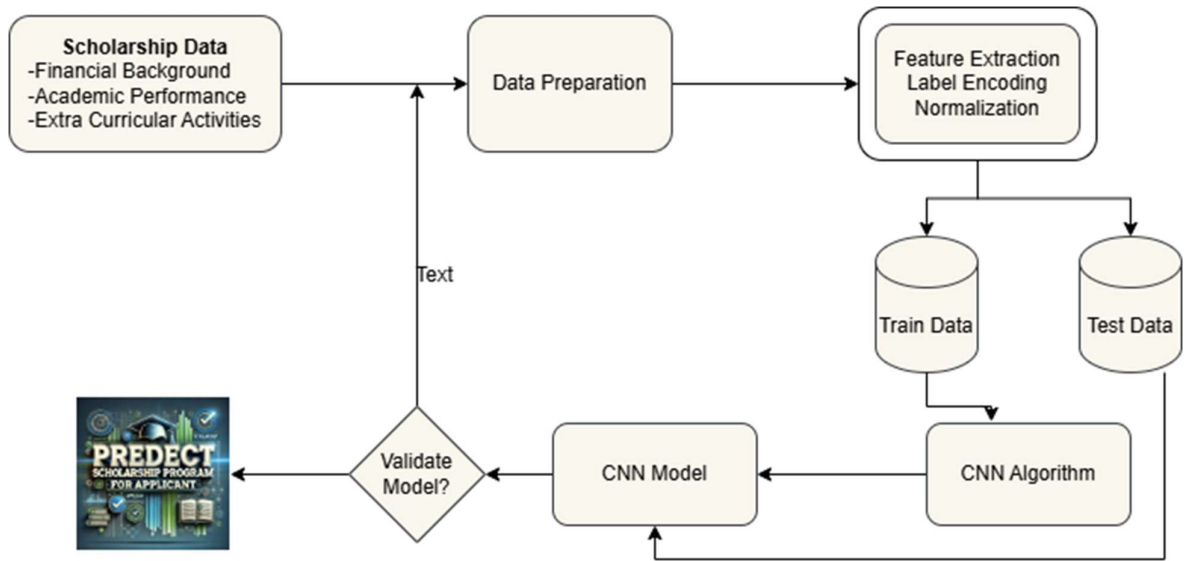
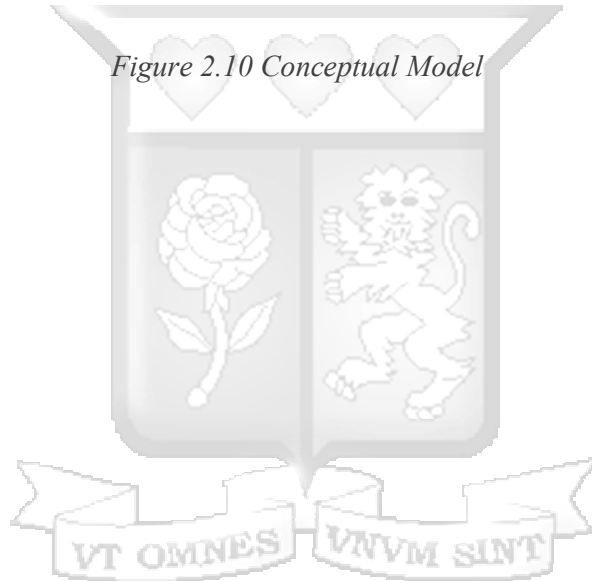


Figure 2.10 Conceptual Model



Chapter 3: Research Methodology

3.1 Introduction

Research methodology refers to the systematic, theoretical analysis of the methods applied to a particular field of study. It involves the strategies, techniques, and procedures used to collect, analyse, and interpret data for a research study (Bhaskar & Manjuladevi, 2016). This chapter outlines the research design, philosophy, population, sampling techniques, data collection methods, data analysis strategies, and the system development approach. Moreover, considerations for research quality, reliability, and ethical issues are discussed to ensure that the research is conducted with rigor and integrity.

3.2 Research Design and Philosophy

Research design refers to the overall strategy that researchers use to integrate the different components of the study in a coherent and logical manner (Bouchrika, 2022). For this research, a descriptive research design was employed. Descriptive research seeks to systematically describe a phenomenon, providing detailed observations and insights (Aggarwal & Ranganathan, 2019). In this study, the design is focused on describing how deep learning algorithms can be applied to the scholarship award process. By utilizing historical data from previous scholarship applications, the research develops a comprehensive understanding of the factors influencing scholarship decisions and developed a deep learning tool to streamline this process.

Research philosophy refers to the set of beliefs concerning the nature of knowledge and how it should be acquired (Matta, 2021). This study adopts a pragmatist philosophy, which emphasizes the use of practical methods that deliver useful outcomes. Pragmatism supports the integration of both qualitative and quantitative approaches to solve real-world problems (Kaushik & Walsh, 2019). In the context of this research, the pragmatic approach allows for flexibility in combining data-driven insights from deep learning models with practical knowledge about scholarship award decisions. This dual approach ensures that the findings are not only accurate but also applicable to real-world scholarship allocation processes.

3.3 Population and Sampling

3.3.1 Population

Population in research refers to the entire group of individuals or items that the researcher is interested in studying (Banerjee & Chaudhury, 2010). For this study, the population consists of scholarship applicants from various higher learning institutions. The target population includes students who have applied for scholarships, with their data spanning academic performance, extracurricular activities, financial need, and demographic factors. These data points were sourced from Kaggle a data sharing platform to obtain dataset that represents different student backgrounds and academic achievements.

3.3.2 Sampling

Sampling refers to the process of selecting a representative subset from the larger population for the purpose of conducting the study (Banerjee & Chaudhury, 2010). In this research, a stratified random sampling technique was employed. Stratified random sampling ensures that the sample accurately reflects the characteristics of the population by dividing it into subgroups (strata) based on certain characteristics such as academic performance, financial need, and demographic background. This ensures that all significant subgroups are adequately represented in the sample. The sample, which includes historical data of scholarship applications, was split into a training set (80%) and a testing set (20%) to train and evaluate the deep learning model.

3.4 Data Collection Methods and Analysis

3.4.1 Data Collection

Data collection involves systematically gathering and measuring variables relevant to the study (Sutton & Austin, 2015). This study utilized secondary data from the "INSPIRE Scholarship 2022-23" dataset, sourced from Kaggle. The dataset provides comprehensive scholarship application records and includes variables such as education qualification, gender, community, religion, prior service in the armed forces (ex-servicemen), disability status, sports involvement, annual percentage scores, household income, geographic location (India or elsewhere), and scholarship outcome (1 for awarded, 0 for not awarded).

The dataset is structured in tabular format, with each record representing an applicant. The inclusion of diverse attributes, such as academic performance, demographic details, and

financial status, allows for a multidimensional analysis of scholarship eligibility. All data is anonymized to protect the applicants' privacy and adheres to data protection regulations. The data is provided in CSV format, facilitating seamless integration with the CNN model for training and evaluation. This structured dataset enables the study to effectively explore patterns and predictors of scholarship allocation while addressing the objectives of fairness and accuracy.

3.4.2 Data Analysis

Data analysis refers to the process of systematically applying statistical and logical techniques to describe and evaluate data (Sutton & Austin, 2015). The analysis in this study involved using the trained CNN model to predict scholarship eligibility based on the input data. Statistical techniques such as cross-validation were employed to ensure that the model is robust and capable of generalizing to new, unseen data. The analysis also involved exploring relationships between the input features and the scholarship award decision to gain deeper insights into the factors that influence scholarship allocation.

3.5 Research Quality and Reliability

Research quality refers to the degree to which the study is able to produce valid and reliable results. Reliability refers to the consistency of the research findings if the study were to be replicated (Leung, 2015). To ensure the quality and reliability of this research, several measures will be implemented:

- i). **Data accuracy:** The data used in this study was sourced from reliable institutional databases, ensuring its integrity.
- ii). **Model validation:** The CNN model underwent 5-fold cross-validation to assess its generalizability to new data. Each fold involved training the model on 80% of the data while testing it on the remaining 20%. The results consistently showed high accuracy (around 90%) but highlighted significant challenges in predicting the minority class (scholarship winners, class 1). While the model achieved near-perfect recall for class 1 (indicating it detected most selected applicants), the precision and F1-score for class 1 were low, ranging from 0.10 to 0.12 across folds. This indicates the model was prone to false positives, misclassifying non-selected applicants as selected. Despite this, the model's performance was stable across all folds, confirming its generalizability. However, the class imbalance issue remains prominent. Future improvements should focus on

enhancing the model's ability to balance precision and recall, potentially through advanced techniques like SMOTE or class weighting adjustments.

iii). **Test-retest reliability:** The model will be tested on different subsets of data to ensure that it produces consistent results across different samples.

3.6 Systems Development Methodology

The system development methodology for this research followed the Agile methodology, which emphasizes iterative development, flexibility, and collaboration between the researcher and the supervisor. Agile development is particularly well-suited for projects where requirements evolve over time and continuous feedback is necessary for improving the system. Figure 3.1 shows the agile methodology.

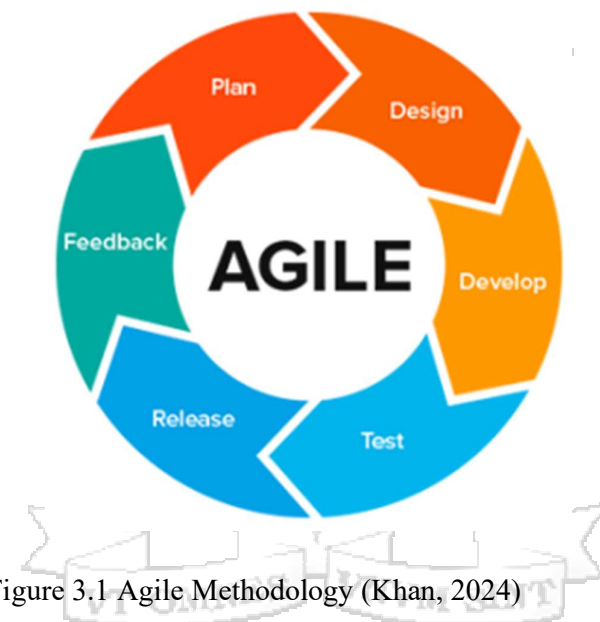


Figure 3.1 Agile Methodology (Khan, 2024)

The key steps involved in the Agile methodology for developing the deep learning tool for scholarship award decisions are as follows:

3.6.1 Planning

The planning phase is the foundation of the Agile methodology. During this phase, the researcher and supervisor collaboratively defined the project scope, set clear objectives, and establish timelines for each sprint. Key requirements for the deep learning tool, such as which student data variables to include (e.g., academic performance, financial need, extracurricular involvement), will be identified. The goals of each sprint, tasks, and deliverables were determined, ensuring that the tool meets the research objectives of fairness, transparency, and scalability in scholarship award decisions.

3.6.2 Design

The design phase involved creating the initial structure and architecture of the deep learning tool. During this phase, both high-level and detailed designs will be developed. This included designing the system's architecture (e.g., how data is processed, stored, and passed through the model) and the structure of the convolutional neural network (CNN) that was used to analyse student data. Unified Modelling Language (UML) diagrams such as Entity-Relationship Diagrams (ERD) and use-case diagrams was used to provide a clear understanding of the relationships between different components of the system. This phase ensured that the system was designed to meet the technical and functional requirements of the project.

3.6.3 Development

The development phase focused on constructing the deep learning tool, a Convolutional Neural Network (CNN) model, using Python and TensorFlow. The CNN architecture included multiple layers tailored to process scoring parameters such as financial need, personal background, and academic performance. Input data was numerically encoded and passed through convolutional layers, which apply filters to extract features and patterns from the data. These features then flowed into pooling layers to reduce dimensionality and computational complexity while retaining essential information. Fully connected layers at the end of the network compiled the extracted features and output a final decision score indicating scholarship eligibility.

The development proceeded incrementally through coding sprints, with tasks such as data pre-processing (cleaning and encoding data), feature extraction (designing appropriate filters for the CNN), and model training. Hyperparameters, including filter size, number of layers, and activation functions, were tuned to optimize model performance. Dropout layers were also incorporated to prevent overfitting, especially when training on large datasets. Regular feedback from the supervisor guided iterative improvements, ensuring the tool aligns with the research objectives. The model underwent continuous testing throughout development, with validation metrics used to monitor performance and identify potential issues early on. This structured approach ensured the CNN architecture effectively processes the scoring parameters and delivers accurate predictions for scholarship allocation.

3.6.4 Testing

In the testing phase, the deep learning tool was rigorously tested to evaluate its performance and accuracy in predicting scholarship award decisions. This phase involved both functional and non-functional testing. Functional testing assessed whether the model accurately classifies scholarship eligibility based on the input data, while non-functional testing evaluated the system's efficiency, scalability, and reliability. The tool was tested using historical data to ensure it generalizes well to new, unseen data. Testing metrics such as accuracy, precision, recall, and F1-score were calculated to determine the model's effectiveness in predicting scholarship outcomes.

3.6.5 Deployment

After the development and testing phases, the deployment phase involved setting up the deep learning tool in a real-world environment. The model was deployed on Heroku, a cloud-based or server infrastructure to allow for scalability and ease of access. The deployment phase also included configuring the user interface for higher education institutions, allowing administrators to input student data and receive scholarship award recommendations. The researcher and supervisor monitored the initial deployment closely to ensure that the system functions correctly and addresses any issues that arise during this phase.

3.6.6 Review and Feedback

The review and feedback phase is a crucial part of the Agile methodology, ensuring continuous improvement. After the initial deployment, feedback from stakeholders, including university administrators and scholarship committees, will be gathered to assess the tool's usability and effectiveness. This feedback will be used to refine the tool in subsequent iterations, making improvements based on real-world performance and user experience. This iterative approach will ensure that the tool remains relevant and aligned with the needs of the institutions using it for scholarship award decisions.

3.6.7 Maintenance and Updates

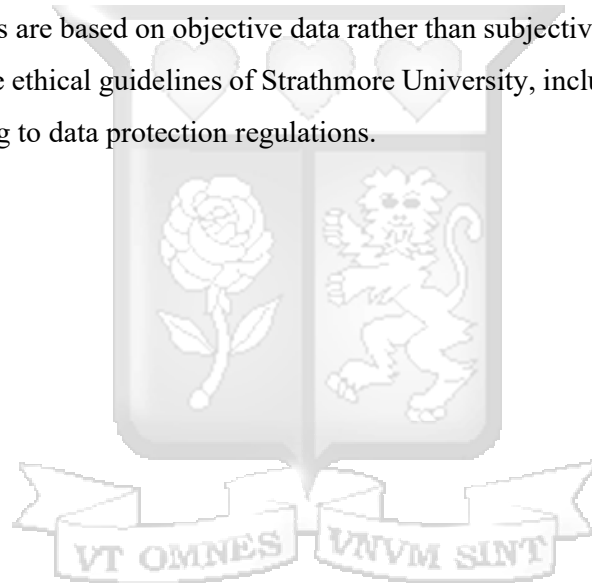
Once the tool is deployed and in use, the maintenance and update phase will ensure that the system continues to operate smoothly over time. Regular updates will be made to the tool to address any issues, improve performance, and incorporate new features based on evolving requirements. This phase ensures the longevity and reliability of the system, making it adaptable to changing data and scholarship criteria as institutions continue to use it in future scholarship allocation cycles.

3.7 Utilization and Dissemination of Research Results

The findings from this research will be disseminated through Strathmore University academic publication and presentations at conferences related to AI, machine learning, and education. Additionally, the developed deep learning tool will be shared with higher education institutions interested in improving their scholarship award processes. The tool will also be made available in relevant open-source repositories, allowing other researchers to replicate and build upon the work.

3.8 Ethical Considerations and Issues

Ethical considerations played a crucial role in this research, particularly regarding data privacy and fairness. Student data used in the model was anonymized to protect the privacy of scholarship applicants. Additionally, the deep learning model was designed to minimize bias, ensuring that decisions are based on objective data rather than subjective factors. The research also complied with the ethical guidelines of Strathmore University, including obtaining ethical clearance and adhering to data protection regulations.



Chapter 4: System Analysis and Design

4.1 Introduction

The development of a deep learning-based scholarship award decision tool involves a detailed analysis and design phase to ensure it meets functional and non-functional requirements. The proposed tool utilizes a Convolutional Neural Network (CNN) to predict scholarship eligibility based on input parameters such as academic performance, financial need, and demographic data. The system will streamline the application, evaluation, and decision-making processes while addressing biases and inefficiencies in traditional methods. This chapter outlines the system's requirements, architecture, and design, including functional specifications, use cases, database schemas, and wireframes.

4.2 Requirement Specifications

The requirement specifications define the functional and non-functional aspects of the proposed scholarship award decision tool. These requirements serve as a blueprint for system development, ensuring that the tool meets its intended purpose of streamlining and automating scholarship allocation processes. Functional requirements outline the specific tasks the system must perform, such as data pre-processing, eligibility prediction, and report generation, while non-functional requirements specify performance, scalability, and security criteria. Together, these requirements provide a comprehensive foundation for designing and implementing a reliable, efficient, and user-friendly system.

4.2.1 Functional Requirements

The functional requirements define the specific capabilities of the scholarship allocation tool:

- i). The system shall provide an online application form for students to input their details.
- ii). The system shall allow administrators to upload existing datasets of scholarship applications in CSV format.
- iii). The system shall predict scholarship eligibility based on input features such as academic performance, financial need, and extracurricular involvement.
- iv). The tool shall calculate a decision score and generate a ranked list of eligible candidates.
- v). The system shall recommend tie-breaking criteria based on predefined rules.
- vi). The system shall generate reports summarizing the prediction results, including statistics and visualizations.
- vii). The system shall provide a secure login for administrators and students.

4.2.2 Non-Functional Requirements

i). Performance:

- a) The system shall process a dataset of 1,000,000 records in under 60 seconds.
- b) Response time for user actions shall not exceed 2 seconds.

ii). Scalability:

- a) The system architecture shall support an increasing number of applications without degradation in performance.

iii). Security:

- a) The system shall use data encryption to ensure the confidentiality of user data.
- b) Adherence to data privacy regulations, such as GDPR, shall be maintained.

iv). Usability:

- a) The user interface shall be intuitive and accessible on both desktop and mobile devices.
- b) Tutorials and tooltips shall guide users through key features.

v). Maintainability:

- a) The system shall allow for easy updates to the CNN model and other components.

4.3 System Architecture

The system architecture incorporates multiple modules, including data collection, pre-processing, CNN-based prediction, and report generation. It ensures seamless interaction between the user interface, prediction engine, and database.

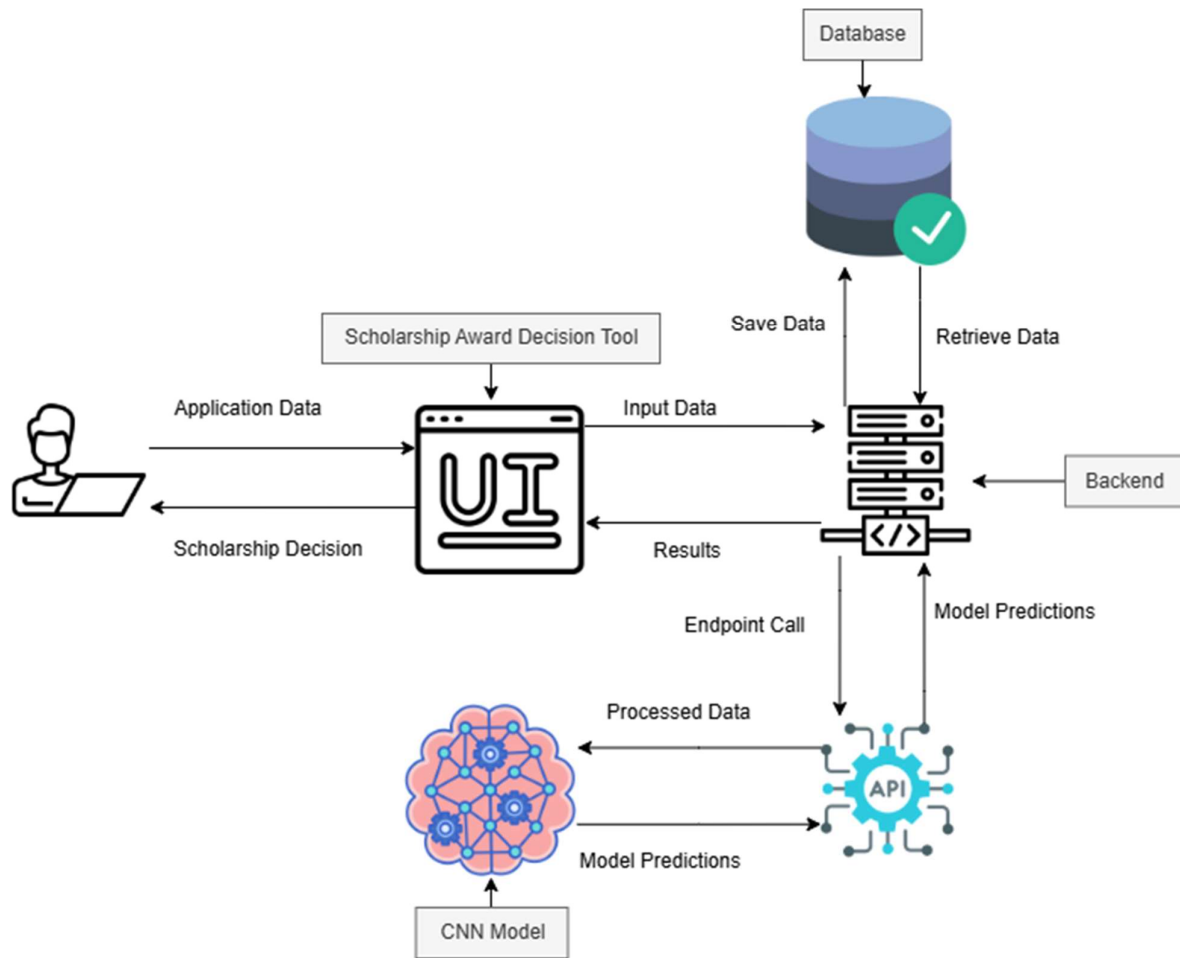


Figure 4.1: System Architecture

4.4 System Design

The design of the scholarship award decision tool follows Object-Oriented Analysis and Design (OOAD) principles, ensuring modularity, reusability, and maintainability. The system's design includes various diagrams to illustrate its structure and behavior:

4.4.1 Use Case Diagram

The use case diagram illustrates the key functionalities of the scholarship award decision tool and how students and the developer interact with the system. Students can fill out and submit scholarship applications. Each use case represents a system feature, ensuring clarity in system behaviour. This diagram provides a high-level overview of user interactions, guiding the development of system functionalities and ensuring that all essential tasks are captured within the system's design.

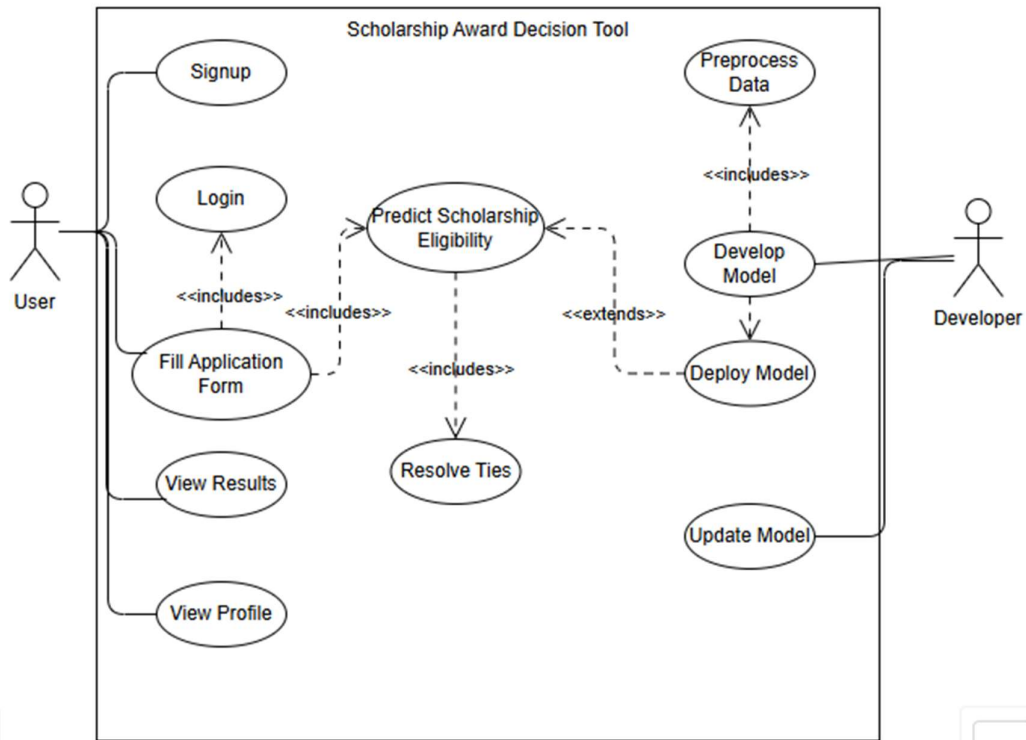


Figure 4.2: Use Case Diagram

3.8.1.1 4.4.1.1 Detailed Use Case Descriptions

Table 4.1 shows the detailed description of use cases in Figure 4.2

Table 4.1: Description of use cases

Use Case	Pre-Conditions	Main Success Scenario	Post Conditions
Fill Application Form	Student logged in	Student submits a completed application form	Form data stored in the database
Upload Scholarship Data	Administrator logged in	System accepts and pre-processes uploaded data	Cleaned data stored in the database
Predict Scholarship Eligibility	Pre-processed data available	CNN model generates predictions	Eligibility results saved in the database
Generate Report	Prediction results available	System compiles results into a report	Report available for download
Resolve Tie	Prediction results with ties exist	System applies tie-breaking criteria	Finalized eligibility rankings saved

4.4.2 Class Diagram

The class diagram defines the structural design of the system, illustrating the relationships between different classes. The main classes include User, Applications and Award Decisions, each containing relevant attributes and methods. The User class represents students, while the Application class holds applicant data. The Award Decision class stores scholarship eligibility decisions made by the Convolutional Neural Network (CNN) model. This diagram ensures that system components are well-structured, promoting code modularity, maintainability, and scalability, while also providing a clear blueprint for developers during implementation.

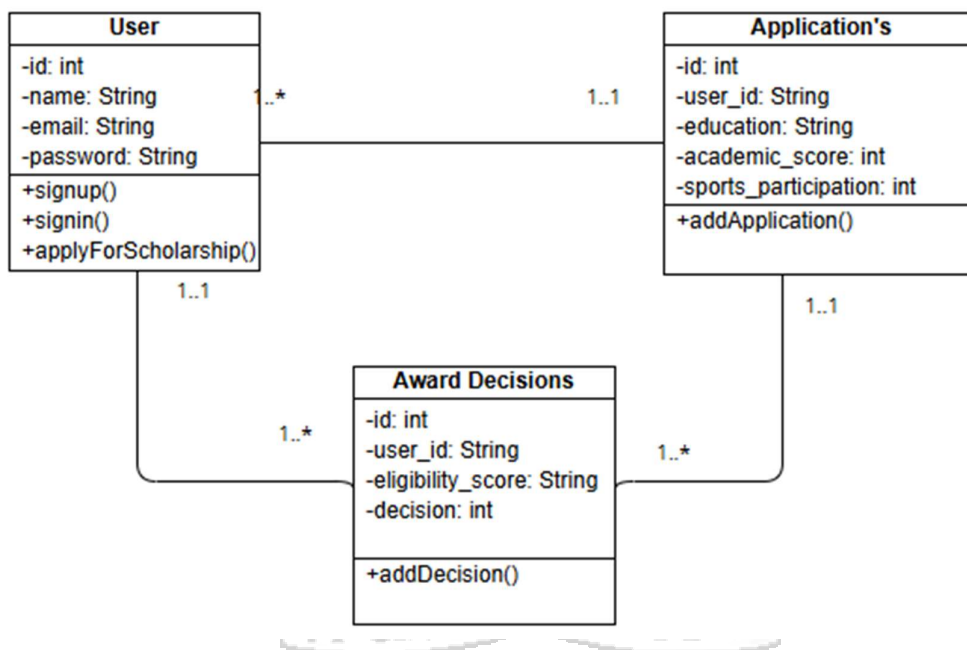


Figure 4.3 Class Diagram

4.4.3 Sequence Diagram

The sequence diagram details the interactions between students and the system throughout key processes. It depicts the sequential flow of actions, such as a student submitting an application and the system processing eligibility through the CNN model. Messages are exchanged between actors and system components to ensure accurate communication. This diagram aids in understanding real-time data flow and system execution order, ensuring that each action follows a logical sequence. It also helps in identifying potential system bottlenecks and improving efficiency in processing scholarship applications.

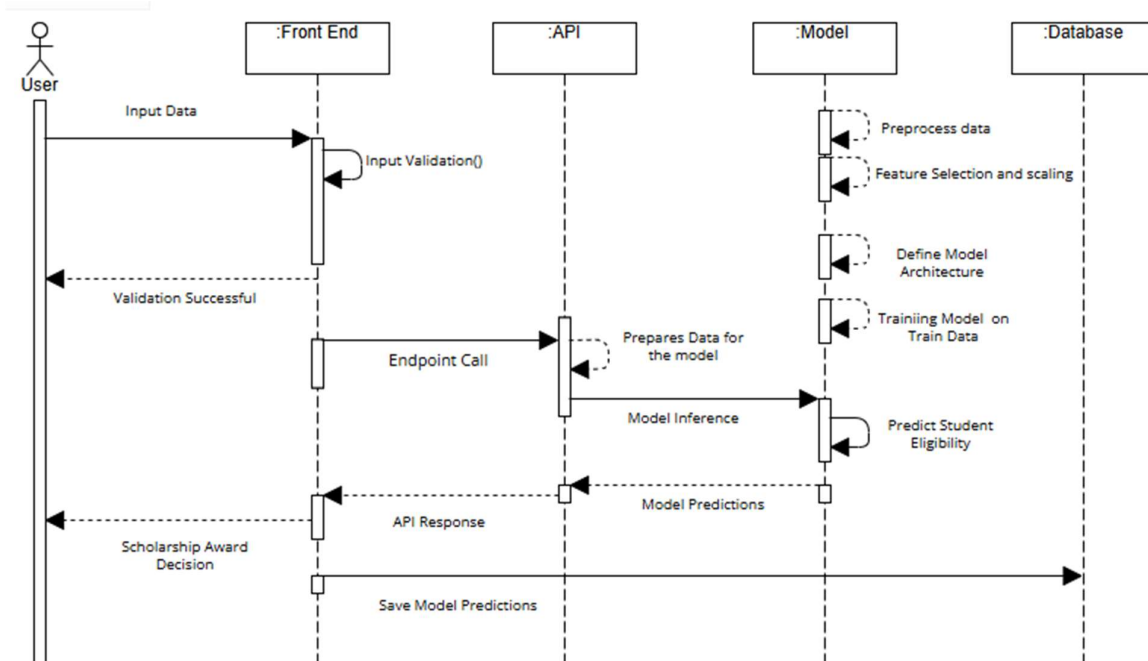


Figure 4.4: Sequence Diagram

4.4.4 Database Schema

The database schema represents the logical structure of the system's database, defining tables, attributes, and relationships. The main tables include Users, Applications, and Award Decisions. The Users table stores login credentials and profile details, while Applications contain student-submitted data. The Award Decision store eligibility results from the CNN model. The schema ensures efficient data organization, enabling seamless data retrieval and updates. It is designed to support scalability, ensuring the system can handle large volumes of scholarship applications while maintaining data integrity and security.

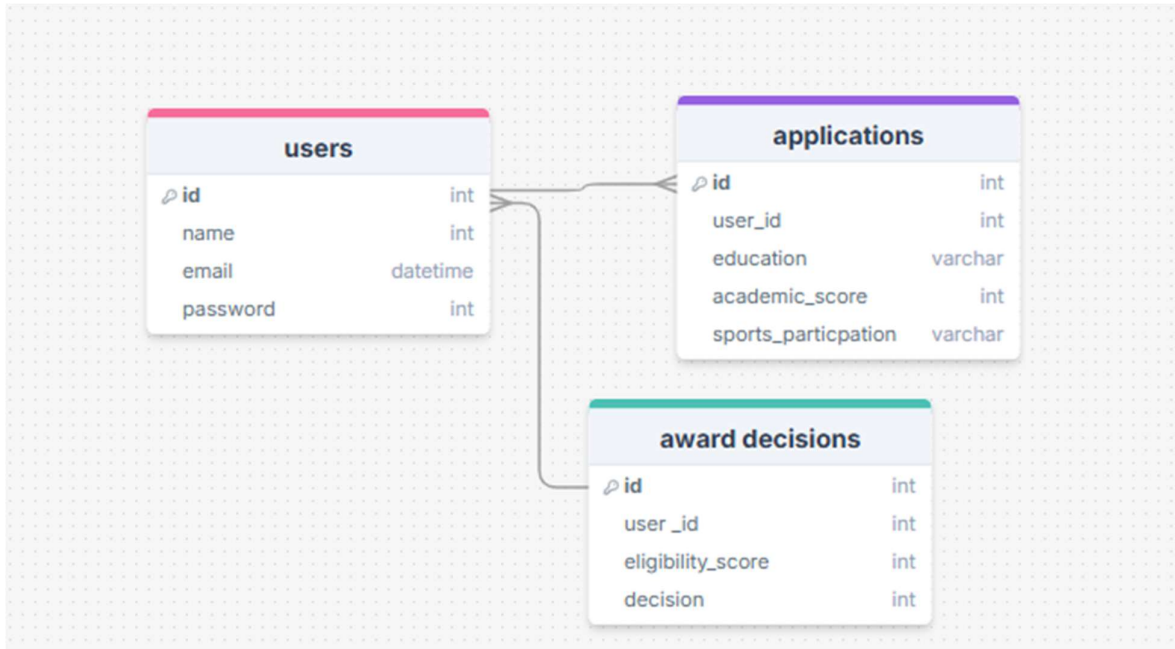


Figure 4.5: Database Schema

4.5 Wireframes

Wireframes serve as the blueprint for the system's user interface, providing a clear and structured visualization of the design and functionality of the application. These wireframes ensure that the user interface aligns with the system's objectives, facilitating user interaction and enhancing the overall user experience. Below are the key wireframes for the Scholarship Award Decision Tool, illustrating the core components and workflows of the system.

4.5.1 Home Page Wireframe

The Home Page Wireframe provides an overview of the system and serves as the entry point for users. It includes navigation links to key sections such as login, registration, and scholarship applications.

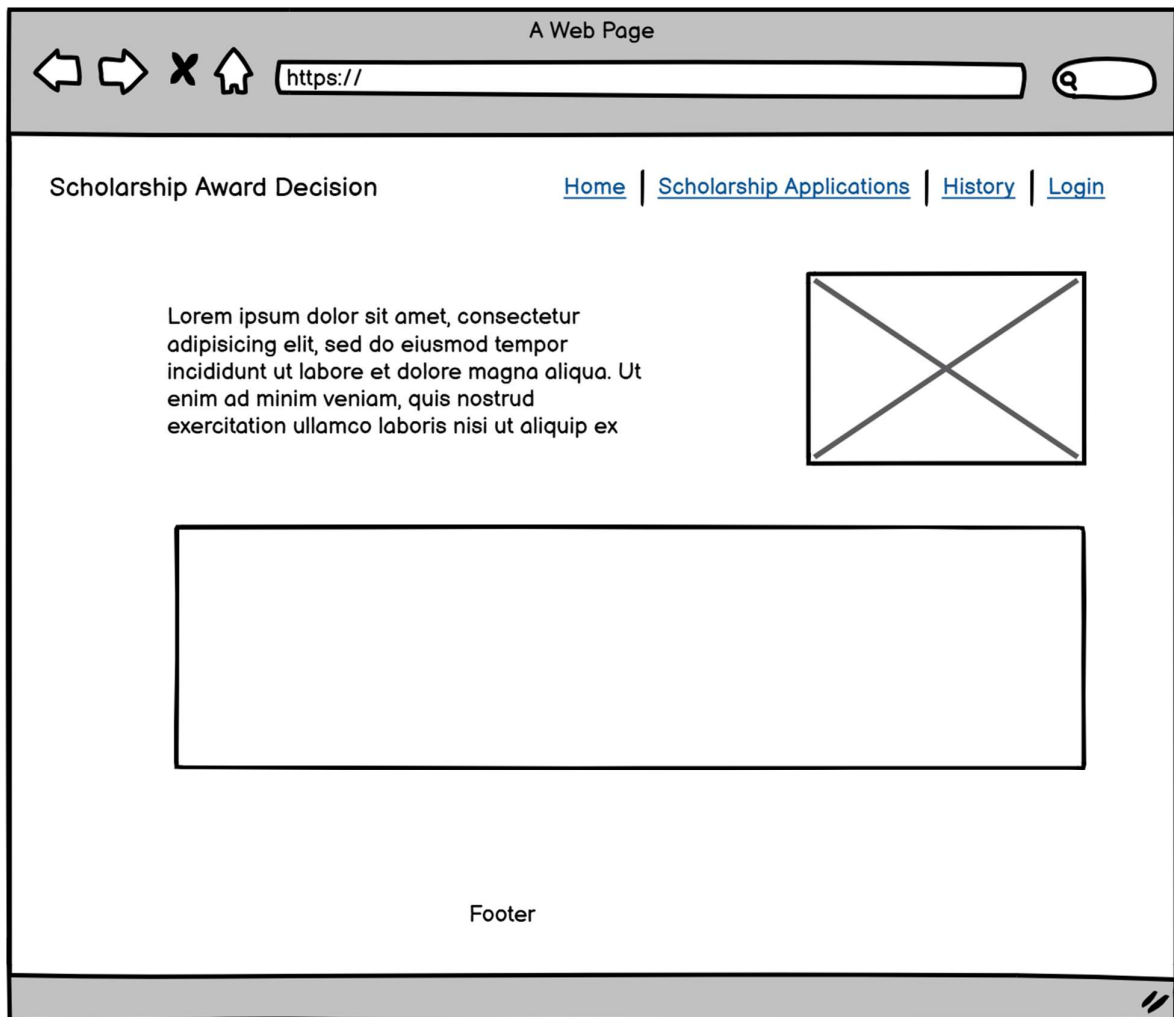


Figure 4.6: Home Page Wireframe

4.5.2 Login Wireframe

The Login Wireframe illustrates the interface where users, including students and administrators, can securely log in to access their respective dashboards.

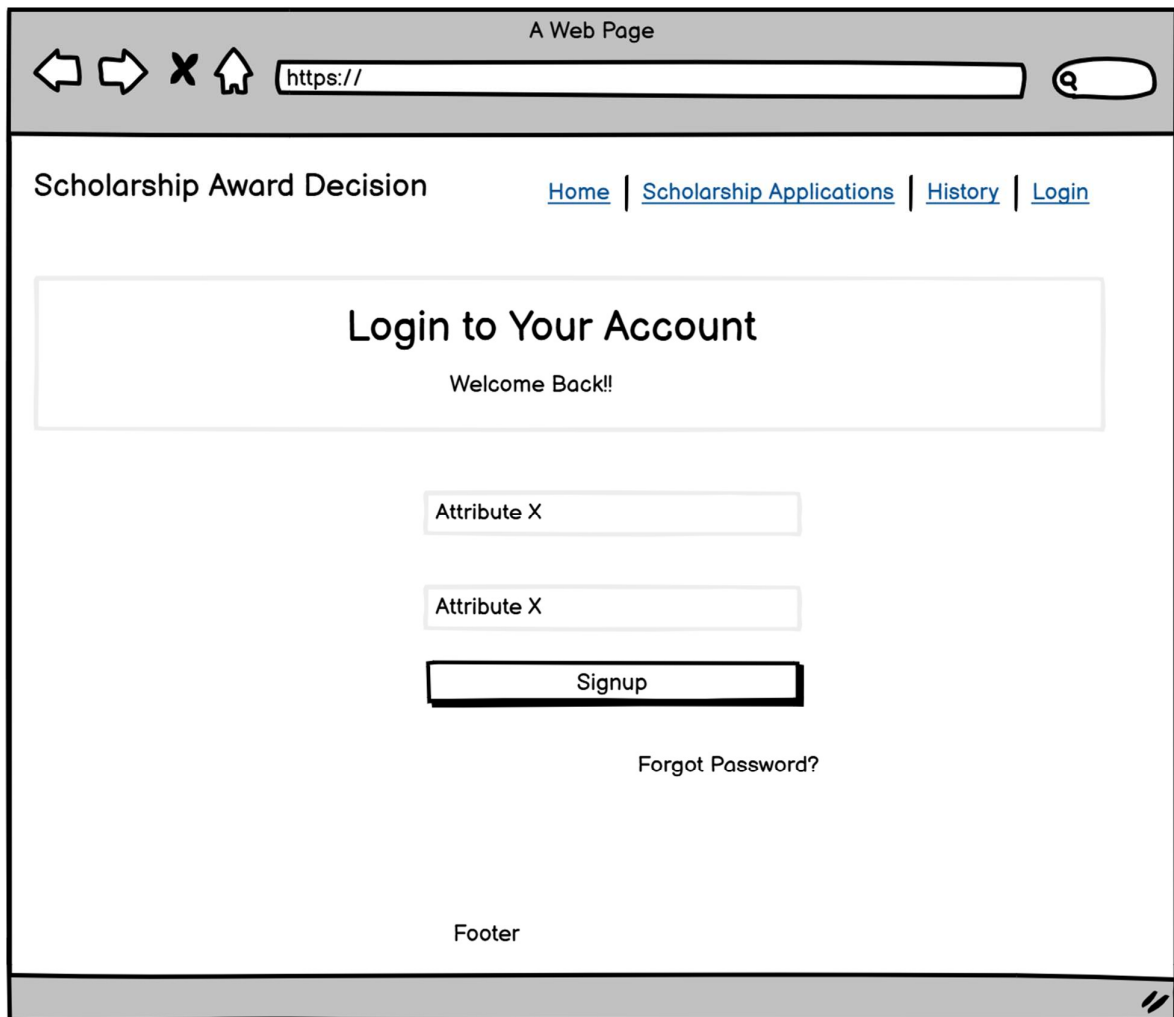


Figure 4.7: Login Wireframe

4.5.3 Scholarship Application Wireframe

The Scholarship Application Wireframe demonstrates the form where students input their personal, academic, and financial details to apply for scholarships.

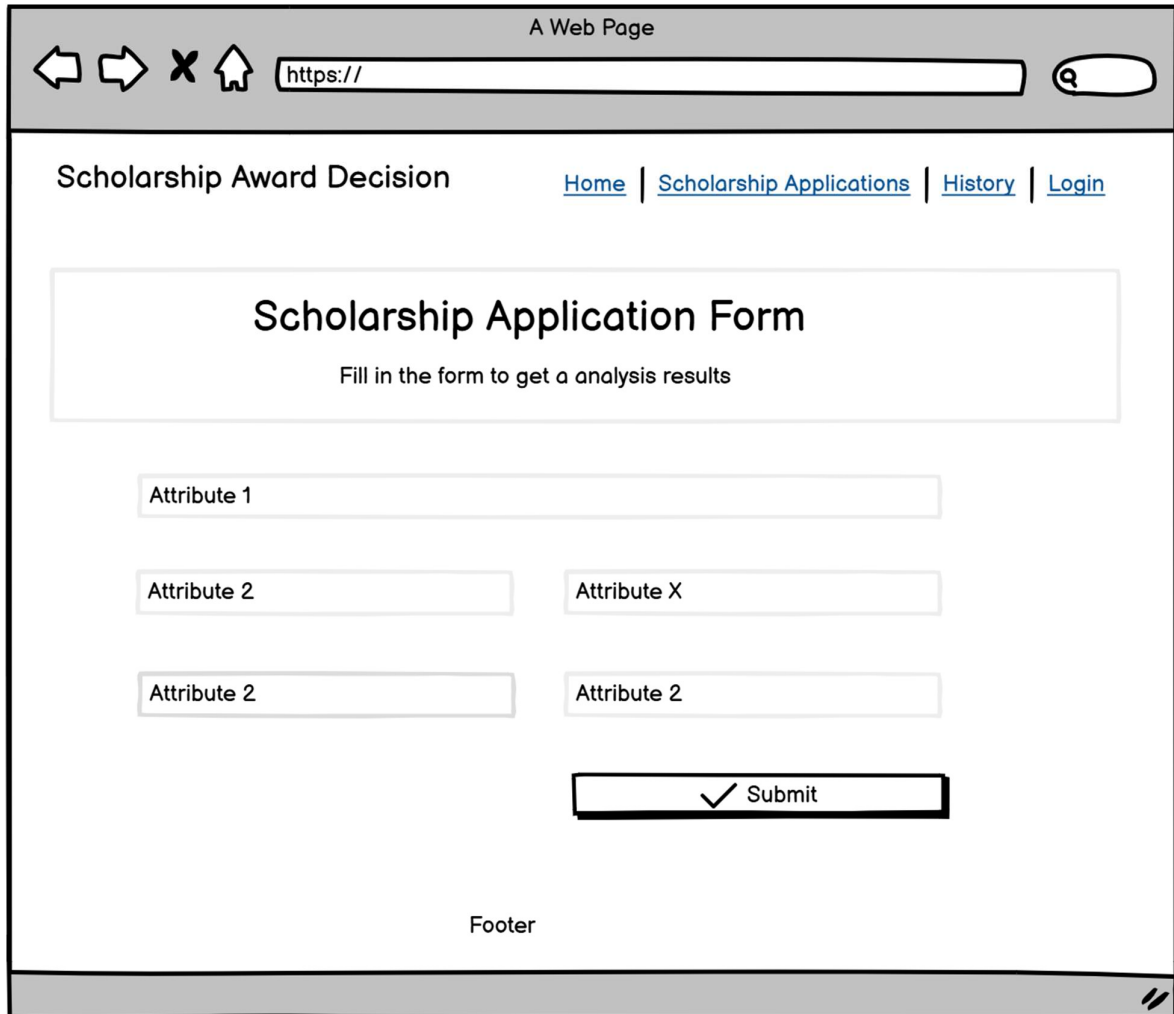


Figure 4.8: Scholarship Application Form

4.5.4 Scholarship Decision Wireframe

The Scholarship Decision Wireframe showcases the interface where administrators review applications and make scholarship award decisions based on the system's recommendations.

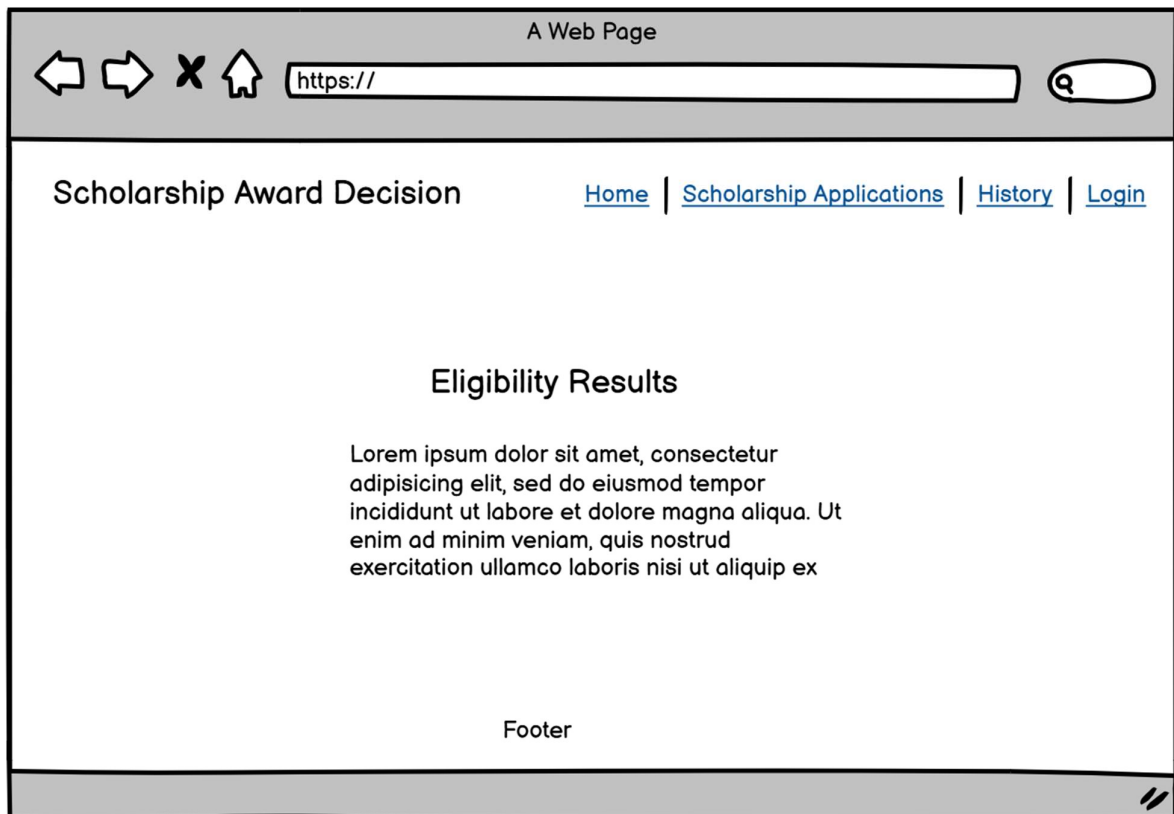


Figure 4.9: *Scholarship Decision Wireframe*

4.5.5 Register Wireframe

The Register Wireframe provides a visual of the registration process, allowing new users to create accounts by entering required personal and contact information.

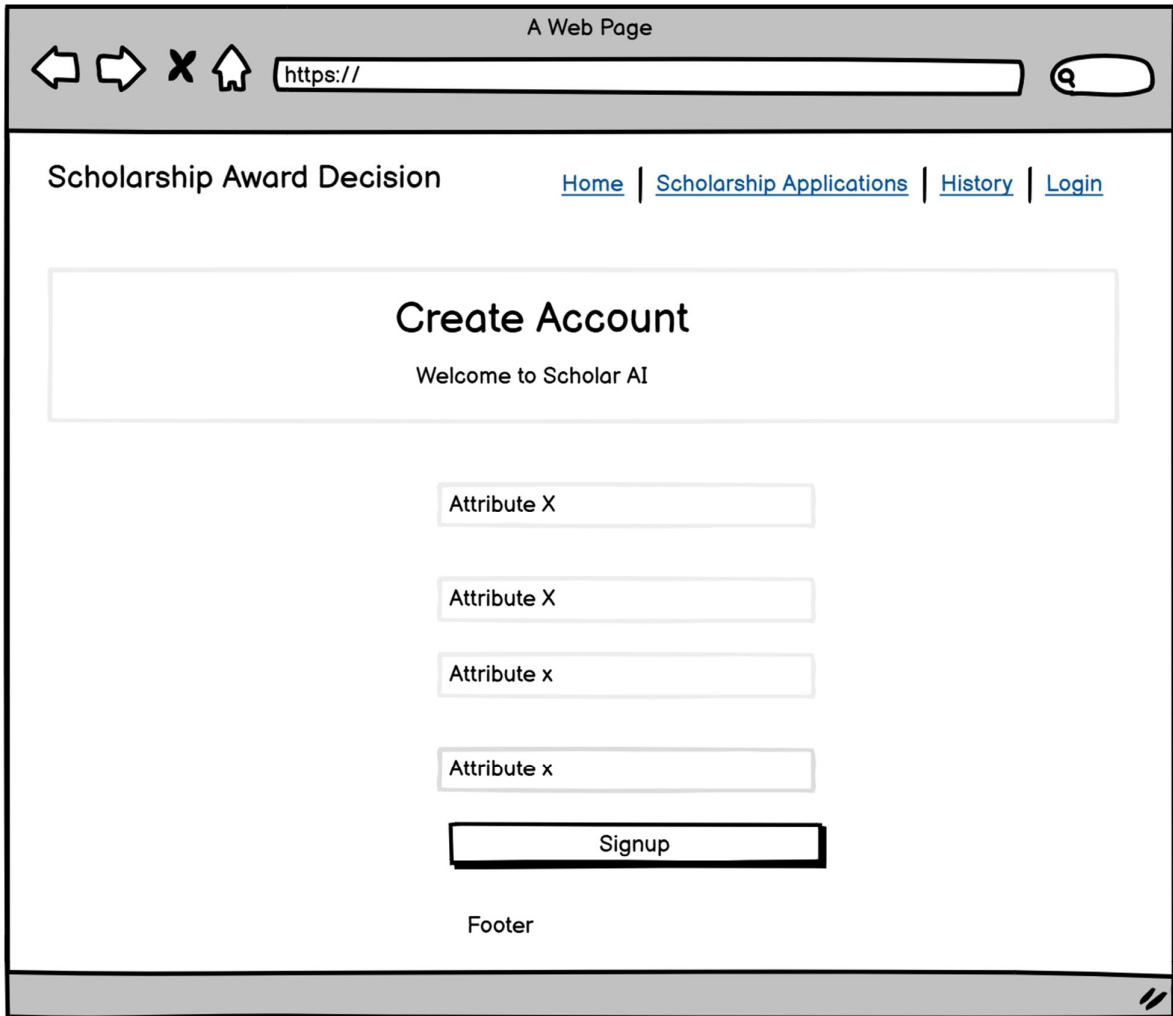


Figure 4.10: Register Wireframe



Chapter 5: System Implementation and Testing

5.1 Introduction

The Scholarship Award Decision Tool was developed to address the challenges of fairness, efficiency, and scalability in the scholarship allocation process. Using a Convolutional Neural Network (CNN) as the predictive engine, the system integrates advanced machine learning techniques with an accessible user interface using Tailwind CSS and HTML. This chapter delves into the detailed implementation of the tool, including the architecture of the CNN model, its integration into a web application, and the testing processes that ensure its reliability and effectiveness. The development environment, data processing techniques, and system components are presented to illustrate how the tool meets its objectives of providing accurate and equitable scholarship decisions.

5.2 Model Components

The Convolutional Neural Network (CNN) model developed for the Scholarship Award Decision Tool is composed of several layers that work together to process and analyse input data, enabling accurate predictions of scholarship eligibility. Each layer serves a specific function, contributing to the overall performance and robustness of the model. Below is a detailed explanation of the model's components, accompanied by code snippets to illustrate their implementation.

5.2.1 Overall CNN Architecture

Figure 5.1 presents the general architecture of a Convolutional Neural Network (CNN), which is a deep learning model designed to process data with grid-like topology. CNNs are particularly well-suited for tasks involving pattern recognition due to their ability to learn hierarchical feature representations directly from raw input data. The architecture begins with an input layer, which accepts structured data in the form of multidimensional arrays. The data is then passed through one or more convolutional layers, where a series of filters are applied to detect local features. Each filter slides across the input data, performing element-wise multiplication and summation, thereby producing a set of feature maps. These maps highlight the presence and spatial location of specific patterns in the input. Following the convolutional layers, activation functions, most commonly ReLU (Rectified Linear Unit), are applied to introduce non-linearity and enable the network to learn complex patterns. Subsequently, pooling layers are used to downsample the feature maps. The most widely used pooling

technique is max pooling, which reduces the spatial dimensions while preserving the most significant features. This also helps minimize overfitting and improve computational efficiency. The output from the final pooling layer is flattened and passed through one or more fully connected layers, which interpret the extracted features and perform high-level reasoning. The network concludes with an output layer, often using a softmax function to generate a probability distribution for classification tasks. This sequential architecture allows the CNN to progressively learn abstract representations and make informed predictions.

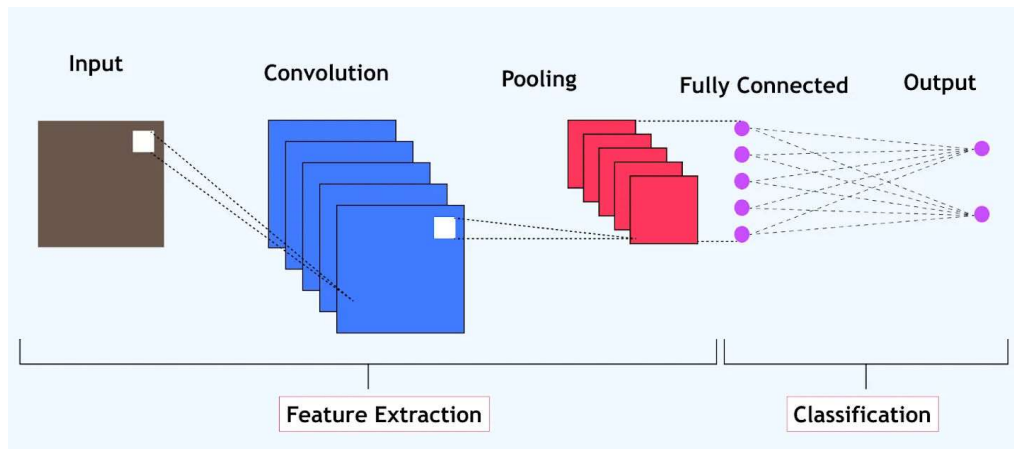


Figure 5.1: CNN Architecture

5.2.2 Convolutional Layers

Convolutional layers are the core of the CNN architecture, designed to extract patterns and features from the input data. The first convolutional layer employs 32 filters with a kernel size of 2. Each filter slides over the input data, detecting local patterns such as correlations between academic performance and financial need. This layer uses the ReLU activation function, which introduces non-linearity to help the model learn complex relationships.

```
Conv1D(filters=32, kernel_size=2, activation='relu', input_shape=(X_train.shape[1], 1)), #
```

Figure 5.2: Convolution Layer

The second convolutional layer expands on this by applying 64 filters with the same kernel size. This layer incorporates padding to preserve the input dimensions, ensuring that no information is lost during the convolution process. The use of ReLU activation continues, enabling the model to capture deeper and more complex features.

```
model.add(Conv1D(filters=64, kernel_size=2, activation='relu', padding='same'))
```

Figure 5.3 Convolution Layers

5.2.3 Pooling Layers

Pooling layers are responsible for reducing the dimensionality of the feature maps generated by the convolutional layers. This reduction decreases computational complexity while retaining the most important information. The model uses a MaxPooling layer with a pool size of 1, which selects the maximum value from each feature map region. By focusing on the most prominent features, the pooling layer enhances the model's ability to generalize, making it more robust when applied to new, unseen data.

```
model.add(MaxPooling1D(pool_size=1))
```

Figure 5.4 Pooling Layers

5.2.4 Dropout Layers

To prevent overfitting, the CNN includes dropout layers that randomly deactivate a fraction of the neurons during training. This forces the network to learn redundant and robust features. The first dropout layer follows the initial convolutional layer and has a dropout rate of 50%. This means that 50% of the neurons are randomly deactivated during training. These dropout layers ensure that the model does not become overly reliant on specific features, improving its ability to generalize across diverse datasets.

```
model.add(Dropout(0.5))
```

Figure 5.5 Dropout Layers

5.2.5 Flattening Layer

After the feature extraction process, the flattening layer converts the multi-dimensional feature maps into a one-dimensional vector. This transformation prepares the data for the fully connected (dense) layers that perform the final classification.

```
model.add(Flatten())
```

Figure 5.6 Flattening Layers

5.2.6 Dense Layers

The dense layers are responsible for combining the extracted features and making predictions about scholarship eligibility. The first dense layer contains 64 neurons and uses the ReLU activation function to learn high-level representations of the input data.

```
model.add(Dense(64, activation='relu'))
```

Figure 5.7 First Dense Layer

The final dense layer serves as the output layer. It consists of a single neuron with a sigmoid activation function, which outputs a probability score between 0 and 1. This score indicates the likelihood of a student being eligible for a scholarship.

```
model.add(Dense(1, activation='sigmoid'))
```

Figure 5.8 Second Dense Layer

5.2.7 Model Compilation

To prepare the model for training, it is compiled with the Adam optimizer, which dynamically adjusts learning rates for efficient training. The binary cross-entropy loss function is used as the model's objective, as the task involves binary classification. Accuracy is selected as the evaluation metric to monitor the model's performance during training and testing.

```
# Compile the model with an optimizer, loss function, and metric  
model.compile(optimizer=Adam(learning_rate=0.001), loss='binary_crossentropy', metrics=['accuracy'])
```

Figure 5.9 Model Compilation

5.3 Scholarship Award Decision Tool

The Scholarship Award Decision Tool integrates the trained CNN model with a web-based application developed using Tailwind CSS and HTML. This tool provides an intuitive user interface for administrators to upload data, process scholarship applications, and view predictions. Below are the key components of the tool.

5.3.1 Landing Page

The landing page serves as the entry point for users, offering navigation options to explore the functionalities of the system. It includes links to analyse application, view history, and access login to the tool.

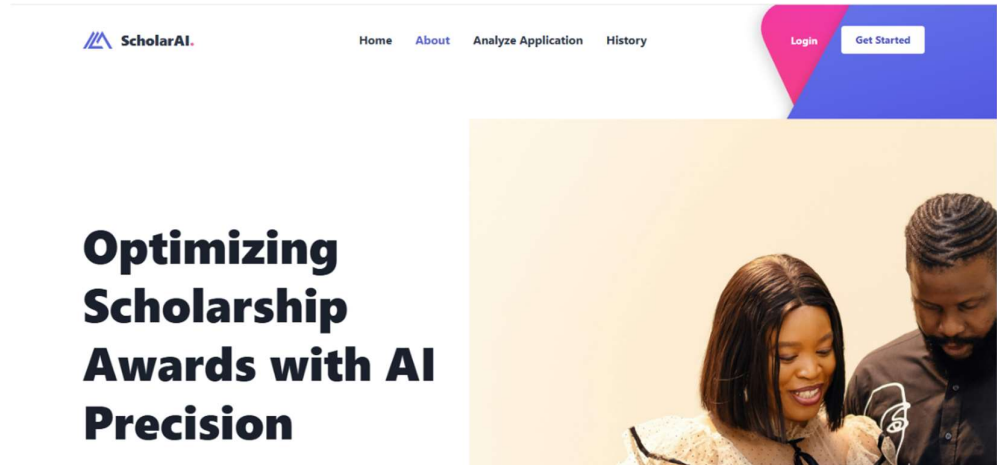


Figure 5.10: Landing Page

5.3.2 Scholarship Application Page

The scholarship application page provides an option for applicants or administrators to complete the scholarship application and view predicted results.

FIRST NAME	LAST NAME	
<input type="text" value="Jane"/>	<input type="text" value="Doe"/>	
EDUCATION QUALIFICATION		
<input type="text" value="Undergraduate"/>		
WHAT IS YOUR ANNUAL INCOME LEVEL?		
<input type="text" value="Upto Upto 223,242 KES"/>		
ANNAL ACADEMIC YEAR SCORE (your last academic year)		
<input type="text" value="60-70"/>		
SELECT YOUR GENDER? (select one option)	DO YOU HAVE DISABILITY?	DO YOU PARTICIPATE IN SPORTS?
<input type="text" value="Male"/>	<input type="text" value="Yes"/>	<input type="text" value="Yes"/>
<input type="button" value="Start Analysis"/>		

Figure 5.11: Scholarship Application Page

5.3.3 Application Results Page

This page presents the predictions generated by the CNN model. It lists whether an applicant is eligible for a scholarship or not. The figures below show the results page for a student who is eligible for a scholarship and the one who is not eligible.

EDUCATION QUALIFICATION

Postgraduate

WHAT IS YOUR ANNUAL INCOME LEVEL?

446,484KES to 892,969 KES

ANNAL ACADEMIC YEAR SCORE (your last academic year)

80-90

SELECT YOUR GENDER? (select one option)	DO YOU HAVE DISABILITY?	DO YOU PARTICIPATE IN SPORTS?
Male <input type="button" value="v"/>	No <input type="button" value="v"/>	No <input type="button" value="v"/>

Student is Not Eligible for Scholarship

EDUCATION QUALIFICATION

Undergraduate

WHAT IS YOUR ANNUAL INCOME LEVEL?

Upto Upto 223,242 KES

ANNUAL ACADEMIC YEAR SCORE (your last academic year)

80-90

SELECT YOUR GENDER? (select one option)

Male

DO YOU HAVE DISABILITY?

No

DO YOU PARTICIPATE IN SPORTS?

Yes

Start Analysis

Student is Eligible for Scholarship



Figure 5.12 Results Page

5.4 System Implementation

5.4.1 Development Environment

- i. The system was developed using the following tools:
- ii. Python: The programming language for model development and integration.
- iii. Google Colab: The environment used for model training, leveraging GPU resources for efficiency.
- iv. Tailwind CSS: The framework for building the web-based user interface.
- v. Flask: Used to create APIs that connect the trained model to the web application.

5.4.2 Scholarship Data Collection

The dataset was sourced from Kaggle, containing variables such as academic performance, financial need, and demographic details. The data was anonymized and pre-processed to ensure compliance with privacy regulations.

5.4.3 Data Pre-processing

Data pre-processing is a critical step in preparing raw data for machine learning models, ensuring it is clean, standardized, and suitable for analysis. The dataset used for this research was sourced from Kaggle and contained various features such as academic performance, financial need, demographic details, and scholarship outcomes. The following pre-processing steps were applied:

5.4.3.1 Loading the Dataset

The dataset was loaded from a specified directory in Excel format using the openpyxl engine. Initial inspections were carried out to check the dataset's structure and ensure data integrity.



```
from google.colab import drive
drive.mount('/content/drive')

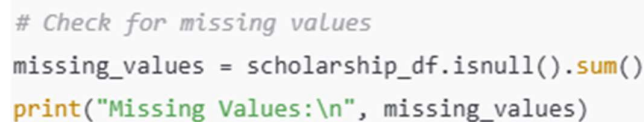
# Define the path to the dataset
dataset_path = "/content/drive/MyDrive/ScholarshipAwardDecisionTool/"
scholarship_df = pd.read_excel(dataset_path + "dataset_combined.xlsx", engine='openpyxl')

# Display the first three rows of the dataset
print(scholarship_df.head(3))
```

Figure 5.13: Load CSV file

5.4.3.2 Handling Missing Values

The dataset was inspected for missing values, which were addressed to maintain data integrity. Features with excessive missing values were dropped, and others were imputed using appropriate statistical methods.



```
# Check for missing values
missing_values = scholarship_df.isnull().sum()
print("Missing Values:\n", missing_values)
```

Figure 5.14 Handling Missing Values

5.4.3.3 Dropping Irrelevant Columns

Certain columns, such as Name, Religion, and Ex-service men, were deemed irrelevant to the model's predictions and were removed from the dataset.

```
# Drop irrelevant columns
scholarship_df = scholarship_df.drop(["Name", "Religion", "Exservice-men"], axis=1)
```

Figure 5.15 Dropping Irrelevant Columns

5.4.4.4 Encoding Categorical Variables

Categorical variables were encoded using OrdinalEncoder for ordinal features and one-hot encoding for binary and multi-class features. Ordinal Encoding for Education Qualification and Income Levels: The Education Qualification and Income columns were ordinally encoded to capture their natural order.

```
def ordinal_encoding(dataframe, col_name, categories, new_col_name):
    enc = OrdinalEncoder(categories=[categories])
    categories = pd.Categorical(dataframe[col_name], categories=categories, ordered=True)
    labels, _ = pd.factorize(categories, sort=True)
    dataframe[new_col_name] = labels
    return dataframe

# Encode Education Qualification
categories = ["Undergraduate", "Diploma", "Postgraduate"]
scholarship_df = ordinal_encoding(scholarship_df, "Education Qualification", categories,

# Encode Income Levels
categories = ["Below 200K", "200K-500K", "500K-1M", "Above 1M"]
scholarship_df = ordinal_encoding(scholarship_df, "Income", categories, "Income_Level")
```

Figure 5.16 Encoding Categorical Variables

5.4.4.5 Separating Features and Target Variable

The target variable (Outcome) was separated from the features, and the remaining columns were prepared for scaling.

```
# Separate features and target variable
y = scholarship_df["Outcome"]
X = scholarship_df.drop("Outcome", axis=1)
```

Figure 5.17: Separating Features and Target Variables

5.4.4.6 Scaling Numerical Features

Numerical features were scaled using StandardScaler to standardize their values, ensuring that no single feature dominated the model's predictions.

```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
numerical_features = X.select_dtypes(include=np.number).columns
X_scaled = scaler.fit_transform(X[numerical_features])

# Create a DataFrame for scaled features
X_scaled = pd.DataFrame(X_scaled, columns=numerical_features, index=X.index)
X = pd.concat([X.drop(columns=numerical_features), X_scaled], axis=1)
```

Figure 5.18: Scaling Numerical Columns

5.4.4.7 Reshaping Data for CNN Input

The data was reshaped into a 3D array to match the input requirements of the CNN.

```
# Reshape for CNN input
X_scaled = X_scaled.values.reshape(X_scaled.shape[0], X_scaled.shape[1], 1)
```

Figure 5.19: Reshaping Data for CNN Input

5.4.4.8 Splitting Data

The dataset was split into 80% training data and 20% test data to facilitate model training and evaluation.

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_s
```

Figure 5.20: Splitting Data

5.4.4 Training Model

The training phase of the Convolutional Neural Network (CNN) was designed to optimize the system's ability to predict scholarship eligibility. This process involved defining the model architecture, compiling it with appropriate configurations, training it on the pre-processed dataset, fine-tuning the parameters, and evaluating its performance. By leveraging the power of CNNs, the model effectively identified patterns in the data, such as the relationships between academic performance, financial need, and demographic details.

5.4.4.1 Model Definition

The CNN architecture was designed to handle multi-dimensional data, extracting meaningful patterns from the input features. The architecture began with two convolutional layers to capture features and patterns, followed by pooling and dropout layers to ensure dimensionality reduction and prevent overfitting. The first convolutional layer consisted of 32 filters with a kernel size of 2, using the ReLU activation function. This layer extracted low-level patterns such as correlations between academic scores and financial data. The second convolutional layer had 64 filters with the same kernel size and padding set to "same" to retain input dimensions, allowing the model to capture more complex patterns. To enhance computational efficiency and reduce overfitting, max-pooling and dropout layers were interspersed after each convolutional layer. The pooling layers summarized the feature maps by selecting the most significant features, while the dropout layers randomly deactivated neurons during training, promoting generalization and robustness. The flattened layer transformed the multi-dimensional feature maps into a one-dimensional vector, preparing the data for the dense layers. Two dense layers followed, with the first layer comprising 64 neurons using ReLU activation to combine high-level features. The final output layer, with a single neuron and sigmoid activation, provided a binary prediction for scholarship eligibility.

5.4.4.2 Model Compilation

After defining the architecture, the model was compiled to prepare it for training. The Adam optimizer was selected for its adaptive learning rate capabilities, ensuring efficient convergence during training. The binary cross-entropy loss function was used due to the binary nature of the prediction task (scholarship awarded or not). Accuracy was chosen as the evaluation metric to monitor the model's performance.

5.4.4.3 Training Process

The CNN model was trained using 80% of the dataset, while 20% was reserved for validation. Training involved running the model for 50 epochs with a batch size of 32. During training, the model updated its weights and biases to minimize the loss function, gradually improving its ability to classify students accurately. The training process monitored both training and validation accuracy and loss. Validation data allowed the detection of potential overfitting or underfitting. By observing the performance on unseen validation data, adjustments could be made to the model's configuration.

```
# Train the model using the training data
history = model.fit(X_train, y_train, epochs=50, batch_size=32, validation_split=0.2, verbose=1)
```

Figure 5.21: Model Training

5.4.4.5 Fine-Tuning

The initial training resulted in an accuracy of 87%. To further improve performance, the model underwent fine-tuning. Adjustments included reducing the learning rate to enable finer updates to the model's parameters, increasing the batch size to stabilize training, and refining the dropout rates to prevent overfitting. During fine-tuning, the model was retrained for an additional 30 epochs with a reduced learning rate and a smaller batch size. These modifications enhanced the model's ability to generalize to new data, resulting in an improved test accuracy of 93%.

5.4.4.6 Model Evaluation

The model's performance was evaluated using the test dataset, which comprised 20% of the original data. Metrics such as loss, accuracy, precision, recall, and F1-score were computed to assess its effectiveness. The model achieved a test accuracy of 93%, indicating its robustness and reliability in predicting scholarship eligibility.

5.4.5 Flask API

To integrate the CNN model with the web application developed using Tailwind CSS, a Flask API was created. The API acted as an intermediary, allowing the trained model to process incoming data from the web interface and return predictions in real time. This architecture ensured flexibility, as the API could handle multiple requests and maintain scalability for future enhancements. The Flask API was implemented with endpoints for receiving data, processing

it through the CNN model, and returning predictions. The Flask library was used to create the server, and the POST method was implemented to handle data submissions.

```
@app.route('/predict', methods=['POST'])
def predict():
    # Parse input JSON data
    input_data = request.json
    input_array = np.array(input_data["features"]).reshape(1, -1, 1)

    # Make predictions
    predictions = model.predict(input_array)
    response = {"prediction": int(np.round(predictions[0][0]))}
    return jsonify(response)

# Run the Flask server
if __name__ == '__main__':
    app.run(debug=True)
```

Figure 5.22: API

5.5 System Testing

System testing was conducted to validate the Scholarship Award Decision Tool's performance, functionality, and scalability. The testing process was essential to ensure the tool met its objectives of accuracy, reliability, and usability in real-world scenarios. Three main areas of focus were model accuracy, functional testing, and scalability testing.

5.5.1 Test on Model Accuracy

The model accuracy tests evaluated the CNN's ability to make correct predictions on unseen data. The test dataset, which comprised 20% of the original data, was used to assess key performance metrics, including accuracy, precision, recall, and F1-score. Accuracy measured the proportion of correct classifications, while precision and recall provided insights into the model's reliability in identifying eligible and ineligible applicants. The F1-score offered a balanced measure of performance by combining precision and recall. The model's predictions were analysed using a classification report and confusion matrix. The classification report provided a comprehensive breakdown of precision, recall, and F1-score, while the confusion matrix visualized the distribution of true positives, true negatives, false positives, and false

negatives. These tools highlighted the model's strengths and identified areas for improvement. The initial training of the CNN model yielded an accuracy of 93%, f1-score of 93% with precision and recall obtaining 95% and 90% respectively.

```

Accuracy: 0.9297
Classification Report:

```

	precision	recall	f1-score	support
0	0.95	0.90	0.93	4858
1	0.91	0.95	0.93	4858
accuracy			0.93	9716
macro avg	0.93	0.93	0.93	9716
weighted avg	0.93	0.93	0.93	9716

Figure 5.23: Model Evaluation

The confusion matrix shows how well the scholarship award decision model performed by comparing its predictions against the actual results. The model correctly identified 4,395 applicants who were not selected and 4,638 applicants who were selected. This means it made accurate decisions in most cases. However, there were some errors. The system mistakenly predicted 463 applicants would be selected when they were not, which means it sometimes gave scholarships to candidates who shouldn't have received them. On the other hand, it missed 220 applicants who should have been selected, failing to award some deserving candidates. Overall, the system's accuracy is quite high, with correct predictions in about 93% of the cases. The higher number of false positives compared to false negatives suggests the system errs more on the side of selecting extra candidates rather than missing qualified ones. This might be a deliberate choice to avoid excluding deserving applicants, but it also means resources might be allocated to some who don't meet the criteria.

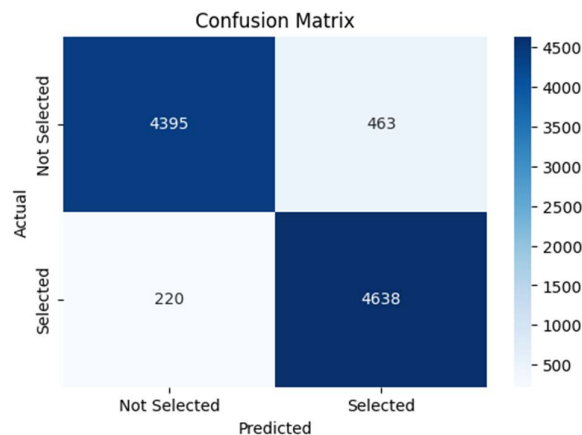


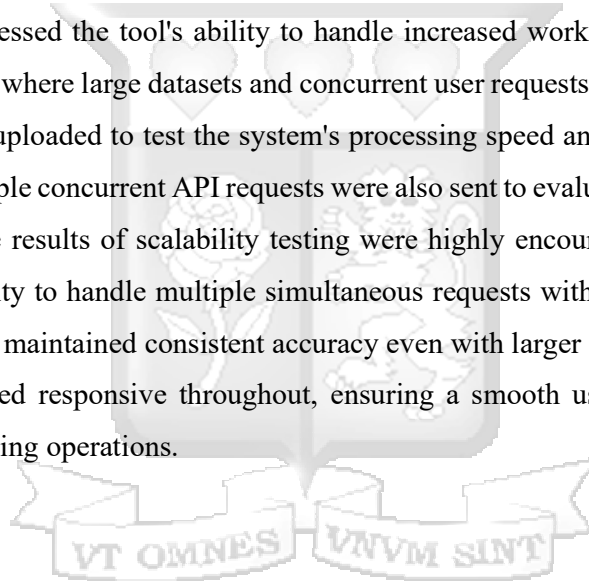
Figure 5.24: Confusion Matrix

5.5.2 Functional Testing

Functional testing focused on verifying the individual components of the tool and their interactions to ensure seamless functionality. The data upload feature was rigorously tested to confirm its ability to handle various file formats, including invalid ones, without crashing. The Flask API was evaluated to ensure accurate communication between the web interface and the CNN model. Predictions generated by the model were checked for correctness and displayed accurately on the interface. Error handling was another critical aspect of functional testing. The system's response to incomplete or incorrect inputs was assessed, and appropriate error messages were verified. These tests confirmed that the tool handled errors gracefully, providing users with clear guidance on corrective actions.

5.5.3 Scalability Testing

Scalability testing assessed the tool's ability to handle increased workloads, simulating real-world usage scenarios where large datasets and concurrent user requests might be encountered. Larger datasets were uploaded to test the system's processing speed and accuracy under high load conditions. Multiple concurrent API requests were also sent to evaluate the responsiveness of the Flask API. The results of scalability testing were highly encouraging. The Flask API demonstrated the ability to handle multiple simultaneous requests without significant delays, while the CNN model maintained consistent accuracy even with larger datasets. The Tailwind CSS interface remained responsive throughout, ensuring a smooth user experience despite extensive data processing operations.



Chapter 6: Discussions

6.1 Introduction

The Scholarship Award Decision Tool was developed to address the significant challenges in traditional scholarship allocation processes. By leveraging deep learning and modern web technologies, this study sought to create a scalable, accurate, and fair solution to manage scholarship decisions. The implementation of a Convolutional Neural Network (CNN) model allowed for comprehensive analysis of diverse applicant data, integrating academic performance, financial need, and demographic factors. This chapter reflects on the findings of the research, contextualizes them within the existing literature, and discusses the broader implications of the tool's development. The discussion highlights the challenges of traditional scholarship allocation processes, compares the CNN-based approach to existing machine learning models, and evaluates the effectiveness of the developed system in addressing equity and scalability concerns.

6.2 Review of Study Objectives

The research was centred on four primary objectives. Throughout the study, each objective was addressed in detail, with the literature review and current findings guiding the fulfilment of the first two objectives. These were subsequently followed by the development and testing of a deep learning tool designed to support decision-making in scholarship awards. The detailed description of each is as follows:

6.2.1 Challenges in the Current Scholarship Allocation Processes

Traditional scholarship allocation processes are fraught with inefficiencies, biases, and scalability issues, as evidenced by both the literature and the findings of this study. Ahmad and Bakar (2020) noted that manual evaluation systems rely heavily on subjective judgments, making them prone to inconsistencies and biases. These systems often fail to provide a holistic evaluation of applicants, focusing disproportionately on standardized metrics like test scores and grades, which may not accurately reflect a student's potential or socio-economic context. Bias is a particularly pervasive issue. Farias-Ruiz et al. (2023) observed that scholarships are often awarded disproportionately to students from privileged backgrounds, who have greater access to resources and preparatory tools. This inequity is compounded using standardized test scores as a primary evaluation metric, which tends to favour students from well-resourced schools (Littler, 2018). The subjective nature of manual processes also introduces unintended biases, as evaluators may be influenced by personal preferences or unconscious stereotypes. Another significant challenge is the inability of traditional systems to scale effectively. With

the increasing volume and diversity of scholarship applications, manual processes struggle to handle the workload efficiently. Jaisri et al. (2024) highlighted that the growing complexity of applications, which include both quantitative and qualitative data, makes it nearly impossible for human evaluators to maintain consistency and fairness across large applicant pools.

6.2.2 Existing Machine Learning Models and Algorithms Used For Scholarship Award Decisions

Machine learning models have been increasingly adopted to address the limitations of traditional scholarship allocation methods. However, these models often fall short in accuracy, scalability, and the ability to integrate holistic evaluation criteria. Ahmad and Abu (2020) introduced ensemble learning techniques, such as Random Forest and Naive Bayes, to predict scholarship eligibility. While these models achieved moderate accuracy (86.9%), they struggled to handle large datasets and failed to consider nuanced factors like socio-economic disparities. Similarly, Aulck et al. (2020b) combined machine learning classifiers with genetic algorithms to optimize scholarship distribution. While their approach increased enrolment yield, it prioritized financial metrics over holistic evaluations, limiting its broader applicability. Other models, such as those developed by Fajardo et al. (2024) and Irfan et al. (2024), relied on simpler algorithms like Naive Bayes and K-Nearest Neighbour's (KNN). These methods were computationally efficient but lacked the ability to process high-dimensional data effectively. Additionally, their reliance on predefined criteria restricted adaptability to diverse contexts, making them unsuitable for large-scale or complex applications. In contrast, the CNN-based approach employed in this research addressed these limitations. The CNN model excelled at extracting patterns from high-dimensional data, integrating diverse features such as academic performance, financial need, and extracurricular involvement. The model achieved a final accuracy of 93%, significantly outperforming existing machine learning models. Furthermore, the scalability of CNNs allowed the tool to handle large datasets without compromising performance, making it suitable for institutions with extensive applicant pools. A key advancement of this research was addressing the interpretability challenge associated with deep learning models. By incorporating transparent pre-processing steps and well-documented feature selection, the study enhanced stakeholder trust in the tool's decision-making process. Additionally, the integration of the CNN model with a user-friendly web interface ensured practical applicability, enabling institutions to adopt the tool with minimal technical expertise. The findings of this research demonstrate the superiority of the CNN-based

approach over traditional machine learning models in terms of accuracy, scalability, and fairness. By leveraging the strengths of deep learning, the developed tool represents a significant step forward in automating and improving scholarship allocation processes.

6.2.3 Deep Learning-Based Scholarship Award Decision Tool Using CNN

The Scholarship Award Decision Tool developed in this research leverages the power of Convolutional Neural Networks (CNNs) to automate the process of evaluating scholarship eligibility. This deep learning-based approach addresses the limitations of traditional methods and existing machine learning models, offering a more accurate, scalable, and fair solution. The CNN model was specifically designed to handle the multi-dimensional nature of the input data, which included features such as academic performance, financial need, demographic details, and extracurricular activities. By employing convolutional layers, the model effectively extracted patterns and relationships within the data, allowing for a holistic evaluation of each applicant. Pooling and dropout layers further enhanced the model's performance by reducing dimensionality and preventing overfitting, ensuring generalizability to new, unseen data.

6.2.4 System Testing

System testing was a critical phase in validating the Scholarship Award Decision Tool's performance, reliability, and readiness for real-world deployment. The testing process focused on three key aspects: the model's accuracy, the system's functionality, and its scalability. Each component of the tool, including the CNN model, Flask API, and Tailwind CSS interface, was rigorously evaluated to ensure seamless integration and robust performance.

6.4.1.1 Model Accuracy

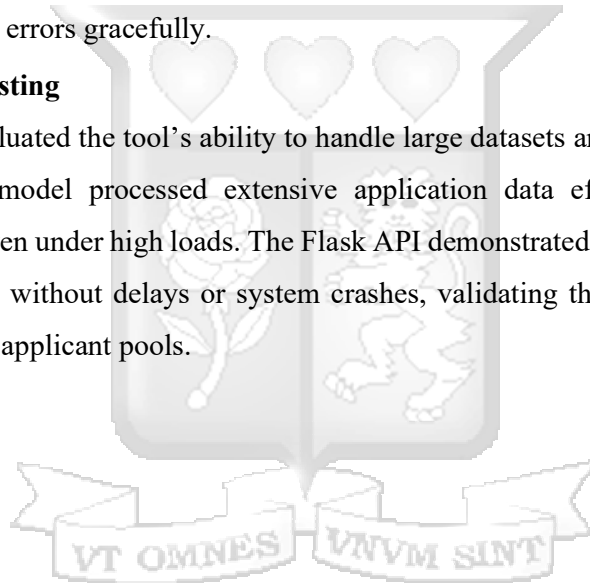
The CNN model achieved an impressive accuracy of 93% during testing, marking a significant improvement over traditional machine learning method. Precision, recall, and F1-score were also calculated to provide a comprehensive assessment of the model's performance. The confusion matrix revealed a high number of true positive and true negative predictions, with minimal false positives and false negatives. This high level of accuracy underscores the reliability of the model in correctly classifying scholarship eligibility. The fine-tuning process played a crucial role in optimizing the model's performance. Adjustments to hyperparameters, such as the learning rate, filter size, and number of layers, significantly improved the model's ability to generalize to unseen data. Cross-validation techniques ensured that the model maintained its accuracy across different datasets, further validating its robustness.

6.4.4.2 Functional Testing

Functional testing ensured that each component of the system operated as intended. The Flask API, which served as the communication layer between the CNN model and the Tailwind CSS interface, was tested for its ability to process incoming requests and return predictions in real time. Various test cases were implemented, including sending valid and invalid data, to verify the API's error-handling capabilities. The Tailwind CSS interface underwent rigorous evaluation to ensure a smooth user experience. Administrators could upload CSV files containing student application data, view predictions, and analyse results without technical complications. The interface provided clear visualizations of outcomes, such as eligibility classifications and confidence scores, making it accessible to non-technical users. The functional testing confirmed that the system was intuitive and user-friendly, with robust mechanisms to handle errors gracefully.

6.4.4.3 Scalability Testing

Scalability testing evaluated the tool's ability to handle large datasets and multiple concurrent requests. The CNN model processed extensive application data efficiently, maintaining consistent accuracy even under high loads. The Flask API demonstrated its capacity to manage simultaneous requests without delays or system crashes, validating the tool's suitability for institutions with large applicant pools.



Chapter 7: Conclusion and Recommendation

7.1 Conclusion

The Scholarship Award Decision Tool developed in this research represents a significant advancement in automating and improving scholarship allocation processes. By leveraging a Convolutional Neural Network (CNN), the tool achieved a remarkable accuracy of 93%, surpassing traditional machine learning models and addressing the inefficiencies and biases inherent in manual allocation methods. The integration of a Flask API and a user-friendly Tailwind CSS interface ensured that the system was practical and accessible for higher education institutions. This study addressed critical challenges in scholarship allocation, including scalability, fairness, and transparency. By incorporating diverse input features such as academic performance, financial need, and demographic data, the tool provided a holistic evaluation of applicants, promoting equity in decision-making. The system's scalability was demonstrated through its ability to process large datasets and handle concurrent user requests, making it suitable for institutions with extensive applicant pools. While the tool exhibited exceptional performance, challenges remain. The interpretability of CNN-based predictions and the incorporation of qualitative data are areas for further enhancement. Nevertheless, this research underscores the transformative potential of deep learning in education, offering a robust, scalable, and equitable solution for scholarship allocation.

In addition to overall accuracy, error analysis through the confusion matrix and classification report provided deeper insight into the model's performance. The confusion matrix showed strong true positive and true negative counts, indicating reliable identification of both selected and non-selected applicants. However, the system produced more false positives than false negatives, suggesting a slight tendency to award scholarships to borderline candidates rather than risk excluding qualified applicants. The classification report further confirmed balanced results with precision and recall values close to 0.91–0.95 across both classes and an F1-score of 0.93, reflecting consistent performance. These findings highlight the model's strength but also point to areas where precision can be improved. Moreover, the CNN's limited interpretability and the lack of qualitative applicant information in decision-making indicate potential directions for future research, such as integrating explainability techniques and richer data sources to enhance fairness and trustworthiness.

7.2 Recommendations

Based on the findings of this research, the following recommendations are proposed for institutions and stakeholders involved in scholarship allocation processes. These recommendations aim to address the challenges identified in traditional and automated methods while leveraging the insights gained from the development of the Scholarship Award Decision Tool.

- i). Institutions should consider transitioning from manual processes to automated, AI-driven systems. By using data-driven models, stakeholders can enhance the accuracy, efficiency, and fairness of scholarship decisions.
- ii). To build trust among administrators, applicants, and the public, institutions should prioritize the use of explainable AI techniques. Providing clear insights into how the model makes decisions will increase confidence in its fairness and reliability.
- iii). Stakeholders should consider incorporating additional variables into their allocation systems, such as leadership roles, community service, or personal statements.
- iv). Institutions should ensure that AI-driven scholarship tools integrate seamlessly with existing administrative and financial systems. This reduces redundancy, streamlines workflows, and enhances decision-making efficiency.
- v). Institutions should track the academic and professional success of scholarship recipients. This feedback can inform adjustments to the system, ensuring that scholarships are awarded to candidates with the greatest potential to succeed.

7.3 Future work

While this research makes significant strides, several areas remain unexplored, offering opportunities for further advancements:

- i). Future research could explore methods for incorporating qualitative factors, such as essays or recommendation letters, into the CNN model. Natural language processing (NLP) techniques could be used to analyse text data, providing a richer evaluation of applicants.
- ii). Future research could focus on developing explainable AI techniques tailored for deep learning models, such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-Agnostic Explanations), to improve transparency and stakeholder trust.

- iii). Future studies could focus on customizing and testing the tool's applicability in different educational systems, cultural contexts, and scholarship types to assess its adaptability and generalizability.
- iv). Future studies could focus on expanding the tool to predict not just eligibility but also long-term outcomes, such as academic success or retention rates, would provide additional value to institutions.
- v). **Mobile Application Development:** A mobile-friendly version of the tool could increase accessibility, particularly in regions with limited desktop infrastructure, enabling administrators and applicants to interact with the system more flexibly.

7.4 Limitations

While the research achieved its objectives, several limitations were identified:

- i). The study focused exclusively on higher education scholarships and did not evaluate the tool's applicability to other educational levels, such as high school or vocational training programs. This narrows the scope and may limit generalizability to other contexts where applicant profiles and criteria differ significantly.
- ii). The dataset used for training and testing relied on secondary data sources, which may not fully capture the complexities, nuances, or evolving characteristics of real-world applicants. This limitation could affect the model's performance when applied in diverse or changing environments.

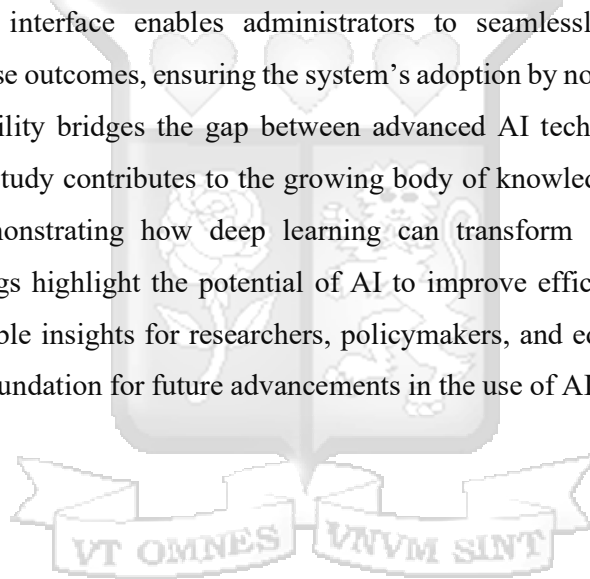
In addition, important ethical concerns arise regarding bias and fairness in the decision outcomes:

- i). The model may inherit biases present in the training data, potentially disadvantaging certain demographic or socioeconomic groups.
- ii). Proxy variables in the data could unintentionally reinforce existing inequalities.
- iii). The system's automated decisions might lack transparency, making it difficult for stakeholders to understand or challenge outcomes.

7.5 Research Contribution

This research makes meaningful contributions to the fields of artificial intelligence, education, and resource allocation by addressing the critical challenges associated with traditional and automated scholarship distribution systems. First, the development of a CNN-based

Scholarship Award Decision Tool achieved a remarkable accuracy of 93%, demonstrating the transformative potential of deep learning for complex, data-driven tasks. The model's ability to process multi-dimensional data effectively positions it as a superior alternative to traditional machine learning techniques, such as Random Forest and Naive Bayes, which struggle with scalability and adaptability. Second, the research introduced a holistic evaluation framework that integrates diverse applicant features, including academic performance, financial need, and demographic factors. This comprehensive approach ensures fairness and inclusivity, addressing systemic biases often inherent in manual and automated systems. By considering the socio-economic context of applicants, the tool promotes equity in scholarship allocation decisions. Third, the integration of the CNN model with a practical, user-friendly web interface showcases the importance of usability in AI applications. The combination of a Flask API and Tailwind CSS-based interface enables administrators to seamlessly upload data, view predictions, and analyse outcomes, ensuring the system's adoption by non-technical users. This emphasis on accessibility bridges the gap between advanced AI techniques and real-world usability. Lastly, the study contributes to the growing body of knowledge on AI applications in education by demonstrating how deep learning can transform scholarship allocation processes. The findings highlight the potential of AI to improve efficiency, scalability, and equity, offering valuable insights for researchers, policymakers, and educational institutions. This research sets a foundation for future advancements in the use of AI for equitable resource distribution.



References

- Adams, J. S. (1965). Inequity In Social Exchange. *Advances in Experimental Social Psychology*, 2(1), 267–299. [https://doi.org/10.1016/S0065-2601\(08\)60108-2](https://doi.org/10.1016/S0065-2601(08)60108-2)
- Aggarwal, R., & Ranganathan, P. (2019). Study designs: Part 2 – descriptive studies. *Perspectives in Clinical Research*, 10(1), 34–36. Ncbi. https://doi.org/10.4103/picr.PICR_154_18
- Ahmad, W. D., & Abu, A. (2020). Ensemble Machine Learning Model for Higher Learning Scholarship Award Decisions. *International Journal of Advanced Computer Science and Applications*, 11(5). <https://doi.org/10.14569/ijacsa.2020.0110540>
- Aulck, L., Nambi, D., & West, J. (2020a). *Increasing Enrollment by Optimizing Scholarship Allocations Using Machine Learning and Genetic Algorithms*. https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_44.pdf
- Aulck, L., Nambi, D., & West, J. (2020b). *Increasing Enrollment by Optimizing Scholarship Allocations Using Machine Learning and Genetic Algorithms*. International Educational Data Mining Society; International Educational Data Mining Society. e-mail: admin@educationaldatamining.org; Web site: <http://www.educationaldatamining.org>. <https://eric.ed.gov/?id=ED608000>
- Banerjee, A., & Chaudhury, S. (2010). Statistics without tears: Populations and Samples. *Industrial Psychiatry Journal*, 19(1), 60–65.
- Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decision Analytics Journal*, 3(100071), 100071. <https://doi.org/10.1016/j.dajour.2022.100071>

- Bansal, M., Nanda, M., & Husain, Md. N. (2021, January 1). *Security and privacy Aspects for Internet of Things (IoT)*. IEEE Xplore.
<https://doi.org/10.1109/ICICT50816.2021.9358665>
- Becker, G. S. (1994). *The Economics of Discrimination*. University of Chicago Press.
- Bhaskar, S. B., & Manjuladevi, M. (2016). Methodology for research II. *Indian Journal of Anaesthesia*, 60(9), 646–651. NCBI. <https://doi.org/10.4103/0019-5049.190620>
- Blaug, M. (1987). *The Economics of Education and the Education of an Economist*.
- Bouchrika, I. (2022). *Types of research design: Perspective and methodological approaches «Guide 2 research*. Research.com. <https://research.com/research/types-of-research-design>
- Bre, F., & Gimenez, J. M. (2024). *Prediction of wind pressure coefficients on building surfaces using Artificial Neural Networks*. Doi.org.
https://www.researchgate.net/publication/321259051_Prediction_of_wind_pressure_coefficients_on_building_surfaces_using_Artificial_Neural_Networks
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/a:1010933404324>
- Bremermann, H. J. (1958). *The Evolution of Intelligence*.
- Budyanto, G., & Suryadi, D. (2023). Model Prediksi dengan Pembelajaran Mesin dalam Pemberian Program Beasiswa kepada Calon Mahasiswa Baru Program S1 di Perguruan Tinggi Swasta. *Jurnal Rekayasa Sistem Industri*, 12(2), 187–200.
<https://doi.org/10.26593/jrsi.v12i2.6595.187-200>
- Chai, S. S., Cheah, W. L., Goh, K. L., Chang, Y. H. R., Sim, K. Y., & Chin, K. O. (2021). A Multilayer Perceptron Neural Network Model to Classify Hypertension in Adolescents Using Anthropometric Measurements: A Cross-Sectional Study in

Sarawak, Malaysia. *Computational and Mathematical Methods in Medicine*, 2021, e2794888. <https://doi.org/10.1155/2021/2794888>

Chaudhari, D., Khadke, R., Sarode, S., Mahajan, S., & Waghulde, L. (2022). Student Scholarship Prediction using Machine Learning Algorithms. *International Journal of Research Publication and Reviews Journal Homepage: Www.ijrpr.com*, 3, 993–996. <https://ijrpr.com/uploads/V3ISSUE3/IJRPR2988.pdf>

Cosentino, C., Fortson, J., Liuzzi, S., Harris, A., & Blair, R. (2019). Can scholarships provide equitable access to high-quality university education? Evidence from the Mastercard Foundation Scholars Program. *International Journal of Educational Development*, 71, 102089. <https://doi.org/10.1016/j.ijedudev.2019.102089>

Delima, A. J. P. (2019). Predicting Scholarship Grants Using Data Mining Techniques. *International Journal of Machine Learning and Computing*, 9(4), 513–519. <https://doi.org/10.18178/ijmlc.2019.9.4.834>

Fajardo, R. C. A., Yara, F. B., Ardeña, R. F., Hernandez, M. K. L., & T. Arroyo, J. C. (2024). A Data-Driven Approach in Predicting Scholarship Grants of a Local Government Unit in the Philippines Using Machine Learning. *International Journal of Engineering Trends and Technology*, 72(6), 74–81. <https://doi.org/10.14445/22315381/ijett-v72i6p108>

Farias-Ruiz, C., Byrd, T., MacLaughlin, E. J., & Hall, R. G. (2023). Evaluation of Scholarship Motivators and Barriers for Non-Tenure-Track Faculty in a Department of Pharmacy Practice. *Pharmacy*, 11(1), 31–31. <https://doi.org/10.3390/pharmacy11010031>

Fix, E. (1985). *Discriminatory Analysis*.

- Fraser, A. (1957). *Simulation of Genetic Systems by Automatic Digital Computers II. Effects of Linkage on Rates of Advance Under Selection*. 10(4), 492–492.
<https://doi.org/10.1071/bi9570492>
- Grosan, C., & Abraham, A. (2011). Artificial Neural Networks. *Intelligent Systems Reference Library*, 281–323. https://doi.org/10.1007/978-3-642-21004-4_12
- Huang, S. (2018). Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics & Proteomics*, 15(1).
<https://doi.org/10.21873/cgp.20063>
- Irfan, M. A. G., Fahad, M., Taariq, A. M., & Anandraj, Mr. P. (2024). Scholarship Matching System Using Naive Bayes And Random Forest. *International Research Journal of Modernization in Engineering Technology and Science*.
<https://doi.org/10.56726/irjmets53073>
- Jaisri, G., Janajeevaraj, V. R., Nandhini, S., Nareen, S. R., & Rajeswari, S. (2024). EQUITABLE SCHOLARSHIP ALLOCATION: HOLISTIC MACHINE LEARNING APPROACH. In *International Research Journal of Modernization in Engineering Technology and Science @International Research Journal of Modernization in Engineering* (pp. 2582–5208).
https://www.irjmets.com/uploadedfiles/paper//issue_4_april_2024/54937/final/fin_irjmets1714925628.pdf
- Jordan, J., Coates, W., Clarke, S., Runde, D., Fowlkes, E., Kurth, J., & Yarris, L. (2018). The Uphill Battle of Performing Education Scholarship: Barriers Educators and Education Researchers Face. *Western Journal of Emergency Medicine*, 19(3), 619–629.
<https://doi.org/10.5811/westjem.2018.1.36752>

- Kadam, V., Kumar, S., Bongale, A., Wazarkar, S., Kamat, P., & Patil, S. (2021). Enhancing Surface Fault Detection Using Machine Learning for 3D Printed Products. *Applied System Innovation*, 4(2), 34. <https://doi.org/10.3390/asi4020034>
- Katoch, S., Chauhan, S. S., & Kumar, V. (2020). A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80(5). <https://doi.org/10.1007/s11042-020-10139-6>
- Kaushik, V., & Walsh, C. A. (2019). Pragmatism as a Research Paradigm and Its Implications for Social Work Research. *Social Sciences*, 8(9), 1–17. MDPI. <https://doi.org/10.3390/socsci8090255>
- Khan, N. A. (2024, May 27). *Incorporating Threat Modeling Techniques into Agile Software Development Processes*. Medium; Medium. <https://medium.com/@noumanalikhan328/incorporating-threat-modeling-techniques-into-agile-software-development-processes-72d4669f0241>
- Kim, C.-H., & Choi, Y.-B. (2017). How Meritocracy is Defined Today?: Contemporary Aspects of Meritocracy. *Economics & Sociology*, 10(1), 112–121. <https://doi.org/10.14254/2071-789x.2017/10-1/8>
- Langford, J., Zhang, X., Brown, G., Bhattacharya, I., Getoor, L., Zeugmann, T., Zeugmann, T., Todorovski, L., Ting, K. M., Corne, D., Handl, J., Knowles, J., Martínez-Jaramillo, S., Alexandrova-Kabadjova, B., García-Almanza, A. L., Centeno, T. P., Martínez-Jaramillo, S., García-Almanza, A., Alexandrova-Kabadjova, B., & Centeno, T. P. (2011). Ensemble Learning. *Encyclopedia of Machine Learning*, 312–320. https://doi.org/10.1007/978-0-387-30164-8_252
- Leung, L. (2015). Validity, Reliability, and Generalizability in Qualitative Research. *Journal of Family Medicine and Primary Care*, 4(3), 324–327. NCBI. <https://doi.org/10.4103/2249-4863.161306>

- Littler, J. (2017). *Against Meritocracy*. <https://doi.org/10.4324/9781315712802>
- Matta, C. (2021). Philosophical Paradigms in Qualitative Research Methods Education: What is their Pedagogical Role? *Scandinavian Journal of Educational Research*, 66(6), 1–14. Tandfonline. <https://doi.org/10.1080/00313831.2021.1958372>
- Mohammed, A., & Kora, R. (2023). A Comprehensive Review on Ensemble Deep Learning: Opportunities and Challenges. *Journal of King Saud University - Computer and Information Sciences*, 35(2). <https://doi.org/10.1016/j.jksuci.2023.01.014>
- Mohd, S., Khutaija, A., Raziuddin, M., & Faisa, O. (2024). *A REVIEW ON DATA MINING AND MACHINE LEARNING METHODS FOR STUDENT SCHOLARSHIP PREDICTION - ProQuest*. Proquest.com. <https://www.proquest.com/openview/f9833560ab8c6b386085359b8357593c/1?pq-origsite=gscholar&cbl=2028922>
- Mountford-Zimdars, A., & Harrison, N. (2016). *Access to Higher Education*. Taylor & Francis.
- Mumtaz, R., Amin, A., M. Ajmal Khan, Asif, M., Anwar, Z., & Muhammad Jawad Bashir. (2023). Impact of Green Energy Transportation Systems on Urban Air Quality: A Predictive Analysis Using Spatiotemporal Deep Learning Techniques. *Energies*, 16(16), 6087–6087. <https://doi.org/10.3390/en16166087>
- Novac, O.-C., Chirodea, M. C., Novac, C. M., Bizon, N., Oproescu, M., Stan, O. P., & Gordan, C. E. (2022). Analysis of the Application Efficiency of TensorFlow and PyTorch in Convolutional Neural Network. *Sensors*, 22(22), 8872. <https://doi.org/10.3390/s22228872>
- Orduña-Ferreira, F., & Franzoni-Velazquez, A. L. (2024). Scholarships Assignments: An Important Challenge of Resources Allocation. Case Study: Completion Prediction

- System for Mexican Institution. *Smart Innovation, Systems and Technologies*, 475–486. https://doi.org/10.1007/978-981-99-7210-4_44
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Louppe, G., Prettenhofer, P., Weiss, R., Weiss, R. J., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12. <https://doi.org/10.5555/1953048.2078195>
- Petrakova, A., Affenzeller, M., & Merkurjeva, G. (2015). Heterogeneous versus Homogeneous Machine Learning Ensembles. *Information Technology and Management Science*, 18(1). <https://doi.org/10.1515/itms-2015-0021>
- Rossi, A., Suwitno Suwitno, & Wijaya, A. (2022). Decision Support System Scholarship Using Profile Matching. *Tech-e*, 6(1), 28–38. <https://doi.org/10.31253/te.v6i1.1468>
- Sengupta, I., Koner, C., Bhattacharjee, N. K., & Gupta, S. (2022, June 1). *Automated Student Merit Prediction using Machine Learning*. IEEE Xplore. <https://doi.org/10.1109/AIC55036.2022.9848976>
- Sutton, J., & Austin, Z. (2015). Qualitative research: Data collection, analysis, and management. *The Canadian Journal of Hospital Pharmacy*, 68(3), 226–231. NCBI. <https://doi.org/10.4212/cjhp.v68i3.1456>
- Zhang, W., & Gao, F. (2011). An Improvement to Naive Bayes for Text Classification. *Procedia Engineering*, 15, 2160–2164. <https://doi.org/10.1016/j.proeng.2011.08.404>

Appendices

Appendix A: Similarity Report

Final Thesis - Draft 1.pdf			
ORIGINALITY REPORT			
17%	14%	15%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS
PRIMARY SOURCES			
1	Submitted to Strathmore University Student Paper	3%	
2	su-plus.strathmore.edu Internet Source	1%	
3	ijettjournal.org Internet Source	1%	
4	open-innovation-projects.org Internet Source	1%	
5	Sri Krishnan. "Machine learning for biomedical signal analysis", Elsevier BV, 2021 Publication	<1%	
6	fci.stafpu.bu.edu.eg Internet Source	<1%	
7	www.mdpi.com Internet Source	<1%	
8	elar.urfu.ru Internet Source	<1%	
9	link.springer.com Internet Source	<1%	
10	Submitted to Georgia Institute of Technology Main Campus Student Paper	<1%	

11	Submitted to Liberty University Student Paper	<1 %
12	mafiadoc.com Internet Source	<1 %
13	Submitted to Heriot-Watt University Student Paper	<1 %
14	scholar.archive.org Internet Source	<1 %
15	dokumen.pub Internet Source	<1 %
16	Submitted to Coventry University Student Paper	<1 %
17	H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in Healthcare Innovation", CRC Press, 2025 Publication	<1 %
18	Submitted to Queensland University of Technology Student Paper	<1 %
19	psasir.upm.edu.my Internet Source	<1 %
20	M.H. Sangdani, A.R. Tavakolpour-Saleh, A. Lotfavar. "Genetic algorithm-based optimal computed torque control of a vision-based tracker robot: Simulation and experiment", Engineering Applications of Artificial Intelligence, 2018 Publication	<1 %

21	Submitted to IUBH - Internationale Hochschule Bad Honnef-Bonn Student Paper	<1 %
22	www.ajhssr.com Internet Source	<1 %
23	Submitted to Sydney Institute of Technology and Commerce Student Paper	<1 %
24	files.eric.ed.gov Internet Source	<1 %
25	listens.online Internet Source	<1 %
26	Mehdi Ghayoumi. "Generative Adversarial Networks in Practice", CRC Press, 2023 Publication	<1 %
27	Shrikaant Kulkarni, Shashikant Bhandari, Dushyant Varshney, P. William. "Artificial Intelligence for Chemical Sciences - Concepts, Models, and Applications", Apple Academic Press, 2025 Publication	<1 %
28	Submitted to Engineers Australia Student Paper	<1 %
29	Submitted to Applied Business Academy Student Paper	<1 %
30	Submitted to Ramapo College Student Paper	<1 %
31	Submitted to University of Technology, Sydney	<1 %

32	Submitted to La Trobe University Student Paper	<1 %
33	Submitted to th-koeln Student Paper	<1 %
34	mmcalumni.ca Internet Source	<1 %
35	www.uyik.org Internet Source	<1 %
36	assets.researchsquare.com Internet Source	<1 %
37	ijariie.com Internet Source	<1 %
38	Submitted to University of Hertfordshire Student Paper	<1 %
39	providing.blogspot.com Internet Source	<1 %
40	Anurag Tiwari, Manuj Darbari. "Emerging Trends in Computer Science and Its Application - Proceedings of the International Conference on Advances in Emerging Trends in Computer Applications (ICAETC-2023) December 21-22, 2023, Lucknow, India", CRC Press, 2025 Publication	<1 %
41	ijitgeb.org Internet Source	<1 %

42	Malti Bansal, Apoorva Goyal, Apoorva Choudhary. "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning", Decision Analytics Journal, 2022 Publication	<1%
43	Submitted to National School of Business Management NSBM, Sri Lanka Student Paper	<1%
44	Submitted to University of Huddersfield Student Paper	<1%
45	Submitted to University of Maryland, Global Campus Student Paper	<1%
46	journal.unpar.ac.id Internet Source	<1%
47	research-api.cbs.dk Internet Source	<1%
48	Ahmad Kumar, Naveen Kumar Chaudhary, Apoorva S. Shastri, Mangal Singh, Anand J. Kulkarni. "Digital Defence - Harnessing the Power of Artificial Intelligence for Cybersecurity and Digital Forensics", CRC Press, 2025 Publication	<1%
49	Mahdi Hosseinipoor, Armin Mollaei Rudsary, Mohammad Danesh-Yazdi, Zahra Kazempour, Yasna Yeganeh. "Why structural solutions for	<1%

flood control should be adapted to climate change?", Natural Hazards, 2024

Publication

50	www.coursehero.com Internet Source	<1%
51	Submitted to Bournemouth University Student Paper	<1%
52	P.V. Mohanan. "Artificial Intelligence and Biological Sciences", CRC Press, 2025 Publication	<1%
53	Submitted to University of Central England in Birmingham Student Paper	<1%
54	ejournal.uksw.edu Internet Source	<1%
55	studentsrepo.um.edu.my Internet Source	<1%
56	"Encyclopedia of Machine Learning", Springer Science and Business Media LLC, 2010 Publication	<1%
57	Min Ruan. "The application of multimodal AI large model in the green supply chain of energy industry", Energy Informatics, 2024 Publication	<1%
58	Submitted to University of Exeter Student Paper	<1%
59	Submitted to University of Sunderland Student Paper	<1%

www.jurnal.yoctobrain.org

60	Internet Source	<1 %
61	www.slideshare.net Internet Source	<1 %
62	Submitted to AUT University Student Paper	<1 %
63	Jeanne, Du Plessis. "Strategies for Psychiatric Nurses to Facilitate Medication Compliance Among Adults Living with Depression.", University of Johannesburg (South Africa), 2024 Publication	<1 %
64	Submitted to Napier University Student Paper	<1 %
65	Saiyed Salim Sayeed, Hemant Kumar Sharma, Pramod Kumar Yadav, Brijesh Mishra. "Advances in Electronics, Computer, Physical and Chemical Sciences - Proceedings of International Conference on Electronics, Computer, Physical and Chemical Sciences (ICEPCPS 2024), July 19-21, 2024, JNRM Port Blair, India", CRC Press, 2025 Publication	<1 %
66	Vieira, Filipe Vieira Borges. "Hydraulic Stability and Wave Overtopping of Single-Layer Cube Armoured Breakwaters", Universidade do Porto (Portugal), 2024 Publication	<1 %
67	aaltodoc.aalto.fi Internet Source	<1 %

68 aiforsocialgood.ca <1%
Internet Source

69 easychair.org <1%
Internet Source

70 www2.mdpi.com <1%
Internet Source

Exclude quotes Off

Exclude matches < 25 words

Exclude bibliography Off

Appendix B: Ethical Clearance Certification



30th January 2025

Mr Nyambura Thomas,
thomas.nyambura@strathmore.edu

Dear Mr Nyambura,

RE: Deep Learning Tool for Higher Learning Scholarship Award Decisions: A Case of Strathmore University

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2564/25**. The approval period is from **30th January 2025 to 29th January 2026**.

This approval is subject to compliance with the following requirements:

- i. Only approved documents including (informed consents, study instruments, MTA) will be used.
- ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
- iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
- iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
- v. Clearance for the export of biological specimens must be obtained from relevant institutions.
- vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
- vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) <https://research-portal.nacosti.go.ke/> and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC