Using Predictive Analytics to Detect Hypertension with Machine Learning

Varsharani T. Dond ; Poonam N. Kale ; Prapti Gaikwad

[1&2] Assistant Professor, PVG's College of Science and Commerce, Pune,Maharashtra, India.

[3] Student, PVG's College of Science and Commerce, Maharashtra, India.

Email_Id: varshadond14@gmail.com[1], poonamkale2730@gmail.com [2]

ABSTRACT

Hypertension, commonly referred to as high blood pressure, is a leading cause of cardiovascular diseases and contributes significantly to premature deaths worldwide. Despite advances in medical treatments, efforts to control hypertension globally have not been fully effective, especially in low- and middle-income countries (LMICs), where access to healthcare resources is limited. A major challenge in managing hypertension is its often undetectable nature; the condition frequently presents no symptoms, leaving individuals unaware of it until serious health issues arise. However, early detection can have a profound impact by preventing complications and reducing both health risks and the financial burden on healthcare systems.

This research explores the use of machine learning techniques to detect hypertension, leveraging the power of algorithms to analyse large datasets, identify patterns, and make predictions based on various clinical and lifestyle factors. By examining patient data—including blood pressure readings, demographic information, and lifestyle habits—our goal is to develop a model that can predict the onset of hypertension at an early stage. This approach has the potential to improve early detection rates and contribute to better public health outcomes, particularly in resource-limited settings.

The primary objective of this project is to develop a machine learning-based tool for early hypertension detection, aiding healthcare providers in making accurate decisions and taking prompt action. This system could play a crucial role in reducing the global burden of hypertension, particularly in LMICs, by enhancing screening efforts and ensuring individuals at risk receive timely and appropriate care.

Keywords : Hypertension, High blood pressure, Undetectable nature, Machine learning, Predictive model, Clinical data, Machine learning algorithms, Health risks

1. INTRODUCTION

1.1 Hypertension Overview

High blood pressure, or hypertension, is a leading cause of cardiovascular diseases and one of the main contributors to premature death around the world. According to the World Health Organization (WHO) Global Report (2023), nearly 1.5 billion people globally are affected by hypertension, and its prevalence continues to grow, particularly in low- and middle-income countries (LMICs). Despite advancements in medical treatments, only about 54% of individuals with hypertension are diagnosed, 42% receive treatment, and only 21% manage to effectively control their blood pressure.
Hypertension significantly increases the risk of severe health conditions, including heart disease, stroke, and kidney failure. Chronic high blood pressure can cause lasting damage to the heart and blood vessels, leading to potentially fatal complications if not addressed. Factors such as excessive sodium consumption, obesity,

lack of physical activity, and poor eating habits further contribute to the widespread occurrence of hypertension, resulting in regional differences in its prevalence.
For example, while the WHO European Region has experienced a decline in hypertension rates, regions like the WHO Western Pacific and Southeast Asia have seen increases in prevalence.

1.2 The Significance of Early Detection

Hypertension is often called the "silent killer" because it usually has no obvious symptoms until serious health problems occur. Many people live with high blood pressure for years without being diagnosed, leading to delayed treatment and a higher risk of severe complications like heart attacks and strokes. Detecting hypertension early is essential, as it allows for timely intervention that can prevent long-term damage to the heart and blood vessels, ultimately lowering the risk of illness and death.However, the silent nature of hypertension makes it difficult to detect. In many cases, individuals don’t realize they have it until significant damage has already occurred. This issue is particularly challenging in low- and middle-income countries (LMICs), where access to healthcare and regular screenings is limited. As a result, a large number of people go undiagnosed and untreated, further straining healthcare systems and worsening the overall health burden.

1.3 Motivation for the Project

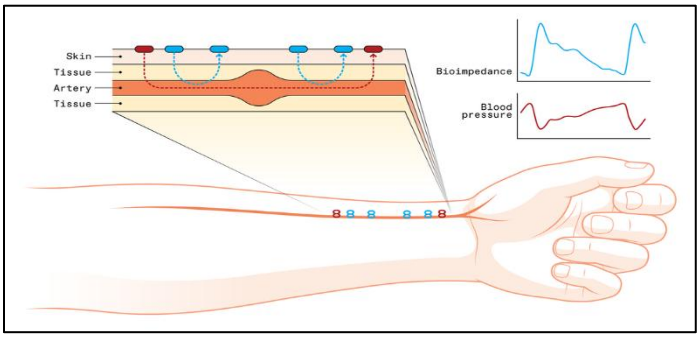
The increasing global rates of hypertension underscore the urgent need for better detection and management strategies. In many parts of the world, especially in low-resource areas, conventional methods of identifying and managing hypertension are insufficient due to a lack of infrastructure, healthcare providers, and public awareness. This gap creates an opportunity for innovation, particularly through machine learning.
Machine learning presents a revolutionary method for detecting hypertension by processing large and complex datasets to identify key trends and make highly accurate predictions. These models can simultaneously evaluate a wide range of factors, including blood pressure, age, gender, personal lifestyle, and family history, which allows for more thorough and streamlined screening processes. This technology can assist healthcare professionals in diagnosing hypertension quickly and accurately, which is especially important in areas with limited access to routine medical visits or diagnostic tools.
By incorporating machine learning into hypertension detection, this project aims to offer a scalable and efficient solution to a major global health issue. The goal is to create a predictive model that can assess the likelihood of hypertension in individuals based on their health and lifestyle data, enabling earlier detection, faster intervention, and a reduction in the overall burden of hypertension worldwide.

2. LITERATURE SURVEY

2.1 Hypertension in a Global Context

Hypertension, or high blood pressure, is one of the biggest health problems in the world, affecting around 1.5 billion people, according to the World Health Organization (WHO, 2023). The number of people with hypertension varies from region to region, with low- and middle-income countries (LMICs) being particularly affected (Kearney et al., 2005). Even though there are medications to treat high blood pressure, the average global blood pressure has stayed mostly the same over the past 40 years, and the number of hypertension cases is increasing in many areas.
In the WHO European Region, efforts to manage and reduce hypertension have been successful, leading to lower rates of the condition. On the other hand, regions like the WHO Western Pacific and Southeast Asia have seen an increase in cases. For example, in the Western Pacific Region, the rate of hypertension grew from 24% in 1990 to 28% in 2019, and in

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930



Southeast Asia, it went up from 29% to 32% during the same period (Peters et al., 2013). These differences are mainly due to factors such as income levels, access to healthcare, and lifestyle choices.

LMICs are especially facing a growing problem with hypertension, driven by factors like rapid urban growth, poor eating habits, and limited access to healthcare (Mendis & Khatib, 2013). The costs related to untreated hypertension are high, leading to more hospital visits, long-term healthcare expenses, and less productivity in the workforce. According to the WHO Global Report (2023), only about 54% of people with hypertension are diagnosed worldwide, and only 21% manage to control their blood pressure properly. This shows that there is a big gap in healthcare, and there's an urgent need for new methods to improve early detection and treatment, especially in places with fewer healthcare resources.

2.2 Factors Contributing to Hypertension

Hypertension, or high blood pressure, is caused by several factors that work together to increase its prevalence. Some of the most common causes include a diet high in salt, lack of exercise, and poor eating habits. These factors are especially noticeable in areas where lifestyles are rapidly changing due to urbanization.

- **High Salt Intake:** Eating too much salt is one of the biggest causes of high blood pressure. Processed foods, fast food, and diets that are heavy on salt are major contributors, especially in cities (Fuchs & Moraes, 2020). A high-salt diet can lead to fluid build-up in the body, which increases blood pressure and puts extra strain on the heart and blood vessels. This is particularly a problem in low- and middle-income countries (LMICs), where diets are becoming more similar to those in Western countries.
- **Lack of Physical Activity:** As cities grow and technology becomes more widespread, people are moving less. Physical inactivity can weaken the heart and blood vessels, making it more likely for someone to develop hypertension. Not exercising regularly can also lead to weight gain, which is another risk factor for high blood pressure (Bhat&Stigum, 2020). In many developing countries, more people are switching from active, physical jobs (like farming) to desk jobs, which makes this problem worse.
- **Poor Diet and Obesity:** Obesity, which often happens because of poor eating and not exercising enough, is a major risk factor for high blood pressure. Diets that are full of unhealthy fats, sugars, and processed foods can lead to weight gain, which raises the chances of developing hypertension (Oparil&Schmieder, 2015). This is becoming more common in LMICs, where it can be harder to find affordable, healthy food.
- **Drinking Alcohol and Smoking:** Drinking too much alcohol and smoking are also key causes of high blood pressure. Both of these habits damage the heart and blood vessels, and they tend to go hand-in-hand in many places around the world. In LMICs, drinking alcohol can be part of cultural traditions, making it harder to reduce alcohol consumption through public health campaigns (Chrysafides&Rajagopalan, 2019).
- **Stress and Economic Factors:** Stress from money problems, job insecurity, and not having access to healthcare can also raise blood pressure. People who live in poverty or in areas with limited resources often deal with chronic stress, which can make high blood pressure worse (Goldstein &

Elad, 2018). In addition, not having access to healthcare means people may not get diagnosed or treated early, making the problem even more serious.

2.3 Machine Learning in Healthcare

Machine learning (ML) is increasingly being applied in healthcare to enhance the diagnosis, treatment, and management of diseases like hypertension. ML can analyze large datasets quickly, identifying patterns and trends that may be difficult for traditional methods to detect. For hypertension, ML algorithms process data such as blood pressure readings, lifestyle factors, and medical history to predict risks and detect early signs of the condition. These models can identify at-risk individuals even before symptoms appear, allowing for early intervention.

ML techniques like decision trees, random forests, and neural networks can improve the accuracy of hypertension detection, leading to more timely diagnoses and better patient outcomes. Additionally, ML models can personalize treatment plans by analyzing individual patient data, ensuring that patients receive the most effective care based on their unique characteristics.

The potential of machine learning in healthcare extends beyond diagnosis to real-time monitoring and continuous risk assessment, making it a valuable tool in managing chronic conditions like hypertension. By improving early detection and treatment, ML can significantly reduce the global burden of hypertension.

2.4 Predictive Models for Hypertension Detection

Several machine learning techniques, such as decision trees, support vector machines (SVM), k-nearest neighbours (KNN), and neural networks, are being used to predict the likelihood of hypertension. These models rely on input data like age, gender, body mass index (BMI), blood pressure readings (systolic and diastolic), lifestyle habits, and family history to assess a person's risk of developing high blood pressure.

Research has shown that these machine learning models can be highly accurate. For instance, a study by Lee et al. (2021) used a random forest algorithm to predict hypertension and achieved an accuracy rate of 85%. The model relied on demographic and clinical data, demonstrating that machine learning can effectively identify individuals at risk for hypertension.

2.5Feature Selection and Data Preprocessing

Feature selection is an essential step in enhancing the performance of machine learning models. By choosing the most important features from a dataset, the model becomes more efficient and can make more accurate predictions. In the case of detecting hypertension, key features such as blood pressure measurements, lifestyle factors, and existing conditions like diabetes are vital for effective predictions (Kearney et al., 2005).

Data preprocessing techniques, including normalization, handling missing values, and adjusting class imbalances in datasets, are also critical to ensure that the model trains on high-quality, clean data. For example, normalization helps to scale data to a consistent range, while handling missing data ensures that valuable information isn't lost during analysis. Addressing class imbalances ensures that the model doesn’t become biased toward the majority class, improving its ability to identify those at risk of hypertension.

2.6 . Challenges and Gaps in Existing Studies

Although many studies show that machine learning can help predict hypertension, there are still some problems. One major issue is that there aren't enough large and diverse datasets from low- and middle-income countries (LMICs), where hypertension is most common (Peters et al., 2013). Most studies use data from richer countries, which makes it hard to apply these models in other regions. Also, many studies do not consider important factors like social and economic conditions, which can affect the risk of getting hypertension.

Another challenge is that most machine learning models are still only tested in labs and have not been widely used in real healthcare settings.

This project will try to solve these problems by developing a machine learning model for hypertension detection that works in places with fewer resources. It will use data that includes social, economic, and lifestyle factors. The goal is to create a model that can be used in real hospitals and clinics to help detect and manage hypertension, especially in LMICs.

3. PROBLEM STATEMENT AND OBJECTIVES

3.1 Problem Statement

Hypertension, or high blood pressure, is a major health issue affecting 1.5 billion people worldwide. It is a leading cause of heart disease, stroke, and early death. Even though there are medicines available to treat hypertension, too many people don't get diagnosed, treated, or have their blood pressure properly controlled. The World Health Organization (WHO) reports that only 54% of people with high blood pressure know they have it, 42% receive treatment, and just 21% manage to control it effectively. These numbers highlight serious gaps in detecting and managing hypertension, especially in low- and middle-income countries (LMICs), where healthcare is often underfunded.

In LMICs, the problem is even worse. There is limited access to healthcare, not enough screening, and people don't always understand how serious high blood pressure can be. Poor diets, lack of exercise, and high salt consumption make the situation worse, and many people are unaware they have hypertension because it doesn't show symptoms until it's too late. If left untreated, it can lead to heart disease, strokes, kidney failure, and a heavy financial burden on both families and health systems.

The most common way to detect hypertension is by regularly checking blood pressure, but this isn't always possible in LMICs because of limited resources. Therefore, there's a need for better, more accessible ways to detect hypertension early. Machine learning could help by using data to predict who is at risk, looking at both medical and lifestyle factors to offer early warnings and help reduce the impact of this condition.

3.2 Objective of the Study

The main goal of this research is to create a machine learning model capable of accurately identifying hypertension based on patient data. By examining various factors, such as blood pressure readings, lifestyle habits (like diet and exercise), personal details (including age and gender), and medical background, this model intends to support the early detection of hypertension, especially in areas with limited healthcare resources.

The specific goals of this study are:

1. **Create a Predictive Machine Learning Model**: Develop a machine learning model that uses both medical and lifestyle information to predict the chances of a patient having hypertension. This model will be trained on a variety of data, including both people with and without hypertension.

2. **Improve Early Detection of Hypertension**: Use this model to identify hypertension at an early stage, before symptoms show up. This will allow doctors to intervene sooner and reduce the risk of serious health problems related to high blood pressure.
3. **Increase Healthcare Access in LMICs**: Address the healthcare issues in low- and middle-income countries (LMICs) by creating a model that can be easily used in areas with fewer resources, where regular blood pressure checks and doctor visits are not always available.
4. **Contribute to Hypertension Research**: Help improve existing research by including more factors, like economic and social influences, as well as regional differences, to make the model more accurate and relevant for different groups of people.

By achieving these goals, the study aims to offer a cost-effective and widely accessible solution for detecting hypertension, helping to reduce its impact globally and fight against this "silent killer."

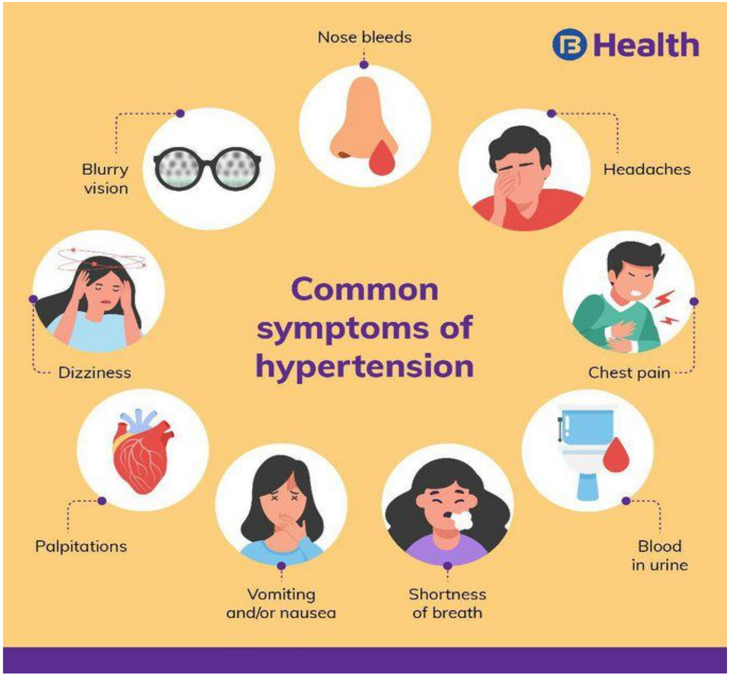
4. DATASET AND DATA COLLECTION

4.1 Description of Data

For this study, we have gathered a detailed dataset that combines patient information from a range of health surveys and data reported by the World Health Organization (WHO). The dataset contains essential variables that are critical for identifying risk factors associated with hypertension. These variables include demographic information, lifestyle habits, and clinical data, all of which are commonly linked to the development of high blood pressure. Below is a breakdown of the key variables:

1. **Demographic Information:**
   - **Age**: The risk of developing hypertension increases with age, making it an important predictor.
   - **Gender**: Men tend to develop hypertension earlier, while women's risk increases after menopause.
   - **Income**: People with higher incomes often have better access to healthcare and healthier lifestyle choices, which can influence the likelihood of developing hypertension.
2. **Blood Pressure Levels:**
   - **Systolic Blood Pressure (SBP)**: This measures the pressure in the arteries when the heart beats. A consistent SBP above 140 mmHg usually indicates hypertension.
   - **Diastolic Blood Pressure (DBP)**: This measures the pressure in the arteries when the heart is at rest between beats. Hypertension is often diagnosed when DBP exceeds 90 mmHg.
3. **Lifestyle Factors:**
   - **Diet**: High sodium intake and poor diet quality are associated with elevated blood pressure. This includes both the consumption of processed foods and the intake of fruits and vegetables.
   - **Physical Activity**: A sedentary lifestyle can contribute to hypertension, while regular exercise is known to help reduce blood pressure.
   - **Smoking**: Smoking is a well-known risk factor for hypertension, as it negatively affects blood vessel health.
   - **Alcohol Consumption**: Drinking alcohol in excess is linked to higher blood pressure levels.
4. **Clinical Data:**
   - **Family History of Hypertension**: A family history of hypertension increases the likelihood of developing the condition, due to genetic factors.

- **Cholesterol Levels**: High cholesterol is often associated with higher blood pressure and cardiovascular issues.
- **Body Mass Index (BMI)**: Obesity significantly contributes to the development of hypertension.
- **Medical History**: Previous conditions like diabetes or kidney disease increase the risk of hypertension.



This dataset includes thousands of records, both from individuals diagnosed with hypertension and those without, providing a balanced view of the factors contributing to hypertension and the patterns observed in hypertensive patients.

4.2 Data Preprocessing

Data Preprocessing for Model Accuracy

The quality and reliability of a machine learning model heavily depend on the preparation of the data. In this project, careful preprocessing was crucial to ensure that the data was clean, consistent, and suitable for training the model. Below is an overview of the preprocessing steps followed in this study:

1. **Handling Missing Values:**

Several variables, especially those related to lifestyle habits such as physical activity and diet, had missing values. These were addressed through different strategies:

- **Imputation with Mean/Median:** For numerical variables like blood pressure and BMI, missing values were replaced with either the mean or median value of the respective variable, depending on the distribution.

- **Imputation with Mode:** For categorical variables like smoking status or gender, missing values were filled with the most frequent value (mode) in that column.
- **Outlier Removal:** Outliers, particularly in blood pressure and cholesterol data, were identified using statistical methods like Z-scores, and either corrected or removed to avoid skewing the results.

2. **Feature Scaling and Normalization:**

Machine learning algorithms often perform better when the data is on a similar scale, especially those based on distance metrics such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM).

- **Min-Max Scaling** was applied to continuous variables such as age, blood pressure, and BMI to bring them within a standard range (0 to 1), ensuring that no one feature dominates the learning process.
- **Standardization** was used for variables with higher variance (like cholesterol levels and income), transforming them into a standardized format with zero mean and unit variance.

3. **Addressing Class Imbalance:**

In datasets where the number of instances in different classes (hypertensive vs non-hypertensive) is unequal, class imbalance can affect model performance. In this project, there were more non-hypertensive cases than hypertensive ones.

- To counter this imbalance, **Synthetic Minority Over-sampling Technique (SMOTE)** was used to artificially generate more hypertensive cases, ensuring the model had a balanced distribution of data from both classes for training.
- Additionally, **under-sampling** of the majority class (non-hypertensive cases) was also considered to maintain the balance and prevent the model from being biased toward the majority class.

5 .MODEL TRAINING PROCESS:

1. **Data Splitting:**
The dataset was split into training (80%) and testing (20%) sets. This ensured the model had enough data to learn from while keeping a separate portion for evaluating its accuracy.

2. **Model Training:**
   - **Logistic Regression:** This model learned the relationship between input features and the probability of hypertension, using metrics like accuracy and confusion matrices to assess performance.
   - **Random Forest:** Multiple decision trees were created using random subsets of features and data, and predictions were averaged from all trees to improve accuracy.
   - **Support Vector Machines (SVM):** The SVM model was trained to find the best boundary between hypertensive and non-hypertensive individuals by adjusting its parameters using kernel functions.
   - **Neural Networks:** If used, this model was trained over multiple iterations, adjusting weights based on the data and optimizing key hyperparameters for improved accuracy.

3. **Model Evaluation:**
After training, models were tested using the evaluation dataset. Metrics like accuracy, precision, recall, and F1-score were calculated to measure performance. Cross-validation was performed to ensure stability across different data subsets. The best-performing model was identified based on these results.

This systematic training approach aimed to create a reliable model for early hypertension detection, improving healthcare outcomes.

6. DISCUSSION

This section presents an analysis of the results obtained from the machine learning models used to predict hypertension, along with the potential implications for healthcare, especially in low- and middle-income countries (LMICs).

1. Key Findings and Interpretation

The results from the machine learning models revealed several important factors that are strongly associated with the likelihood of developing hypertension:

- **Age:** Older individuals are at a higher risk for hypertension due to natural age-related physiological changes and a buildup of other risk factors over time. This highlights the importance of regular blood pressure checks for older adults, who show higher hypertension rates.
- **Body Mass Index (BMI):** Higher BMI, often a sign of obesity, is closely linked to an increased risk of hypertension. Excess weight can cause higher blood volume and increased strain on the heart, leading to elevated blood pressure. This shows the need for programs focused on weight management to prevent hypertension.
- **Physical Activity:** A sedentary lifestyle is a key risk factor for hypertension. On the other hand, regular physical activity has proven benefits in lowering blood pressure. Encouraging exercise as part of public health initiatives can significantly reduce the incidence of hypertension.

These results emphasize the role of both lifestyle choices and demographic factors in the development of hypertension. This insight is especially valuable for LMICs, where limited healthcare resources and lack of awareness contribute to a high burden of hypertension. By focusing on the most important risk factors, such as age, BMI, and physical activity levels, targeted interventions can be developed to effectively prevent and manage hypertension in these regions.

2. Healthcare Implications

The machine learning model developed in this research offers several valuable applications in real-world healthcare settings, especially in areas with limited resources:

- **Improved Screening and Early Detection:** The model can act as an efficient screening tool in regions with limited access to routine health checks, like rural areas. By identifying individuals at risk for hypertension based on data such as age, BMI, and lifestyle factors, healthcare providers can prioritize high-risk patients and implement early interventions.
- **Community Health Programs:** The findings of this study can guide public health campaigns focused on preventing and managing hypertension in communities. Health initiatives can be designed around the key risk factors identified by the model, like diet, exercise, and obesity management. This could lead to tailored community-based interventions, such as educational programs that emphasize the importance of healthy living and regular blood pressure monitoring.
- **System Integration in Healthcare Settings:** The model can be incorporated into existing healthcare systems and electronic health records, enabling continuous patient monitoring. This would allow healthcare providers to track hypertension risk over time, monitor treatment outcomes, and adjust care plans as needed to ensure better management of hypertension.
- **Efficient Resource Allocation:** By identifying high-risk individuals, the model can help healthcare systems distribute resources more efficiently. Targeting interventions to those most likely to develop hypertension ensures that healthcare services are provided where they are needed most, improving healthcare delivery efficiency and patient outcomes.

This approach promises to improve the quality of care and reduce the burden of hypertension in regions with limited healthcare resources.

7 CONCLUSION AND FUTURE WORK

7.1 Conclusion

In conclusion, the machine learning model created to detect hypertension presents significant opportunities for improving public health, especially in resource-constrained settings. The model highlights key risk factors, such as age, BMI, and physical activity, which play a critical role in the development of hypertension. By applying these insights to real-world healthcare systems, particularly in low- and middle-income countries (LMICs), we can create targeted interventions that reduce the impact of hypertension.

The results emphasize the importance of integrating technology into healthcare, alongside community-based initiatives and policy changes, to effectively tackle hypertension. Combining data-driven tools like machine learning with education, outreach, and better healthcare policies can help identify high-risk individuals, promote early detection, and improve management strategies.

The "Detection of Hypertension Using Machine Learning" project has successfully emphasized the importance of early hypertension detection, a condition that significantly impacts public health and cardiovascular well-being. By applying various machine learning algorithms, this study has shown how demographic, lifestyle, and clinical data can be harnessed to predict hypertension risk. The project highlights how machine learning can be a game-changer in detecting hypertension, especially in areas where healthcare services are limited. By adopting these advanced technologies, healthcare systems can prioritize interventions and screenings, leading to better outcomes for at-risk individuals. Furthermore, the findings can guide public health campaigns aimed at increasing awareness and preventing hypertension, ultimately reducing its prevalence and related health complications.

7.2 Future Work

Although this project has made great progress, there are still several areas that can be improved:

1. **Adding More Data:** Future models can be better by including more factors like genetic information, environmental influences, and more details about people's diets. This could help make the predictions more accurate.
2. **Using Data from More People:** Expanding the dataset to include different populations and regions will make the model more reliable for people from different backgrounds. This could also make the model work better in different healthcare systems around the world.
3. **Including Other Health Problems:** This research could be used to detect other heart-related issues, like heart disease or strokes, by using similar machine learning techniques. This would help create a broader approach to heart health.
4. **Community Education and Programs:** The success of this machine learning model also depends on teaching people about hypertension and its risks. There should be programs to raise awareness about early detection. Working with local healthcare providers and communities will help spread this knowledge and encourage healthier lifestyles.

In short, machine learning has great potential to improve the detection and management of hypertension. With more data, better models, and strong community education, we can help reduce the challenges of hypertension, especially in low- and middle-income countries where the disease burden is highest.

8 REFERENCES

1. World Health Organization (WHO). (2023). *Global Report on Hypertension: The Silent Killer*. Retrieved from WHO Website.
2. Kearney, P. M., Whelton, M., Reynolds, K., et al. (2005). Global burden of hypertension: Analysis of worldwide data. *Lancet*, *365*(9455), 217-223. https://doi.org/10.1016/S0140-6736(05)17741-1
3. Peters, S. A., Huxley, R. R., & Woodward, M. (2013). The burden of hypertension and its associated risk factors in the developing world: A global perspective. *European Journal of Preventive Cardiology*, *20*(3), 404-409. https://doi.org/10.1177/2047487312441896