**Dissertation Proposal**

**Student:**

Nabiya Yousaf

**Student ID:**

2312514

**University of Essex**

**Table of content**

**[Introduction:](#_Toc172702509)** [3](#_Toc172702509)

[**Aim of the project:** 3](#_Toc172702510)

[**Research Question:** 3](#_Toc172702511)

[**Significance of the project:** 3](#_Toc172702512)

[**Progress to date:** 5](#_Toc172702513)

[**Task 1: Understanding Heart Diseases:** 5](#_Toc172702514)

[**Task 2: Literature Review:** 5](#_Toc172702515)

[**Task 3: Data Collection:** 5](#_Toc172702516)

[**Planned Work:** 6](#_Toc172702517)

[**Week 1: July 24 - July 30** 6](#_Toc172702518)

[**Week 2: July 31 - August 6** 6](#_Toc172702519)

[**Week 3: July 31 - August 6** 6](#_Toc172702520)

[**Week 4: August 14 - August 20** 7](#_Toc172702521)

[**Week 5: August 21 - August 27** 7](#_Toc172702522)

[**Week 6: August 28 - September 3** 7](#_Toc172702523)

[**Bibliography** 8](#_Toc172702524)

[**Appendices** 9](#_Toc172702525)

[**Appendix 1: Detailed CNN Architectures Explored** 9](#_Toc172702526)

[**Appendix 2: Data Preprocessing Techniques** 9](#_Toc172702527)

[**Appendix 3: Evaluation Metrics Definitions** 10](#_Toc172702528)

# **Introduction:**

Heart disease, which includes a variety of heart-related disorders, continues to be one of the world's top causes of death. According to estimates from the World Health Organization (WHO), cardiovascular diseases (CVDs) cause around 17.9 million deaths a year, or 31% of all fatalities worldwide. For people with heart disease, early identification and treatment are essential to lowering mortality and raising quality of life. But sometimes the accuracy and promptness of traditional diagnosis techniques—which frequently rely on clinical judgment and conventional diagnostic instruments—can be compromised. This is where convolutional neural networks (CNNs), a type of machine learning, come into play and show promise for improving the predictive modeling of cardiac disorders. (See Appendix 1)

Convolutional neural networks represent a family of deep learning algorithms that have become quite popular because of their remarkable ability to recognize images. CNNs are specially made to identify patterns in visual data, unlike standard neural networks, which makes them very useful for interpreting medical pictures like MRIs, ECGs, and X-rays. CNNs are an effective tool in medical diagnostics because of their capacity to automatically and adaptively learn the spatial hierarchies of information from input images.

# **Aim of the project:**

This project's main aim is to use patient data to forecast cardiac problems using CNNs. Improved heart disease prediction models and a dependable tool to aid medical practitioners in early identification are the driving forces behind this research. CNNs can be used to swiftly and accurately examine massive amounts of data, spotting patterns and anomalies that human eyes might miss, as part of the diagnostic process. In addition to ensuring that patients receive prompt and proper medical care, this also expedites the diagnostic process.

# **Research Question:**

* "How effectively can convolutional neural networks predict heart diseases using patient data?" is the main research question of this project. This entails investigating several CNN architectures and methodologies in order to identify the best precise and effective model for the prediction of cardiac disease.
* What is the impact of different data preprocessing techniques on the accuracy of CNN models in predicting heart diseases?
* How does the inclusion of demographic and lifestyle factors (such as age, gender, smoking habits, and exercise frequency) influence the predictive power of CNN models for heart disease?

# **Significance of the project:**

This initiative is important because it has the potential to change the way that cardiovascular healthcare is provided. Despite their effectiveness, traditional diagnostic techniques are frequently constrained by the subjective interpretation of healthcare professionals and the inherent unpredictability of patient presentations. CNNs in particular, which are machine learning models, provide a more objective and consistent method of diagnosis, decreasing the possibility of mistakes and increasing overall diagnostic accuracy. These models may also learn from fresh data and keep improving, keeping them current with the most recent findings and procedures in medicine.

In order to accomplish the objective of employing CNNs for heart disease prediction, this research will adhere to a systematic methodology. The initial phase is gathering an extensive dataset with a range of patient characteristics, including age, gender, blood pressure, cholesterol, and medical background. This information will be gathered from credible medical databases and earlier research, guaranteeing its breadth and accuracy.

Data preparation is the next step after data collection. This is an important stage that includes encoding categorical variables, addressing missing values, and normalizing the data. By preparing the data, one may make sure that the CNN model can learn from it and that the dataset is appropriate for training.

Model selection will be the project's primary focus after data preprocessing. To determine which CNN architecture is best for predicting cardiac illnesses, a number of them will be assessed. To maximize the performance of the model, this entails testing with various network depths, filter sizes, and activation functions.

Next, the preprocessed dataset will be used to train the chosen model. When a CNN is trained, input data is fed into the model, and its weights are adjusted to reduce prediction errors. Iterative in nature, this approach necessitates large computational resources in addition to close supervision to avoid overfitting or under fitting.

The model will be assessed using a variety of measures, including accuracy, precision, recall, and F1-score, following training. These metrics offer a thorough evaluation of the model's functionality, showing both its strong points and potential areas for development. In order to make sure the model is trustworthy and able to generate precise predictions on fresh, untested data, the assessment stage is essential.

The project's last phase involves compiling the reports and analyzing the results. The efficacy of CNNs in predicting heart illnesses will be determined by analyzing the evaluation outcomes of the model. The final report, which will include a description of the research's methodology, findings, and conclusions, will be informed by this study. Furthermore, the study will incorporate suggestions for subsequent investigations, examining possible enhancements and expansions to the present model.

# **Progress to date:**

The best way of doing any kind of project irrespective of its type is dividing it into small parts like milestones or tasks. So for conducting my research I also divide my projects into small tasks in a sequential manner. As I am in the middle of the whole project now in this section I will discuss the tasks I have done till now.

## **Task 1: Understanding Heart Diseases:**

**Status: Completed**

I read medical literature and spoke with medical experts to get a thorough understanding of heart conditions. This basic understanding is essential for determining pertinent aspects and deciphering model results.

## **Task 2: Literature Review:**

**Status: Completed**

I did a thorough literature study on current approaches to heart disease prediction, with a particular emphasis on CNNs and machine learning. This influenced the development of my methods and assisted in identifying gaps in the literature.

## **Task 3: Data Collection:**

**Status: Completed**

I gathered a dataset with a variety of patient characteristics, including medical history, blood pressure, cholesterol, and age and gender. The dataset came from earlier research and reliable medical databases.

# **Planned Work:**

In this section, I will write about the remaining tasks which I have to perform now to complete my research.

## **Week 1: July 24 - July 30**

**Task 4: Data Preprocessing:**

**Status: Remaining**

 **July 24 - July 27: Initial Data Preprocessing**

* Handle missing values, standardize data, and encode categorical variables.
* Ensure the dataset is suitable for training the CNN model.

 **July 28 - July 30: Advanced Data Preprocessing**

* Apply advanced preprocessing techniques to the dataset.
* Finalize the dataset for optimal model training.

## **Week 2: July 31 - August 6**

**Task 5: Model Selection and Initial Training**

**Status: Remaining**

 **July 31 - August 3: Model Selection**

* Evaluate several CNN models and select promising candidates based on their performance in related tasks.

 **August 4 - August 6: Initial Model Training**

* Start training the selected CNN models using the preprocessed dataset.
* Iteratively adjust parameters to optimize model performance.

(See Appendix 3)

## **Week 3: July 31 - August 6**

**Task 6: Continued Model Training and Evaluation Preparation**

**Status: Remaining**

 **August 7 - August 10: Continued Model Training**

* Continue iterative training and parameter tuning.
* Document the training process and preliminary results.

** August 11 - August 13: Prepare for Model Evaluation**

* Finalize the trained models for evaluation.
* Ensure all necessary evaluation metrics are in place.

## **Week 4: August 14 - August 20**

**Task 7: Model Evaluation**

**Status: Remaining**

 **August 14 - August 17: Model Evaluation**

* Evaluate the trained models using metrics such as accuracy, precision, recall, and F1-score.
* Compare the performance of different models to identify the best-performing one.

 **August 18 - August 20: Detailed Evaluation Analysis**

* Analyze the evaluation results to understand the strengths and weaknesses of the models.
* Interpret the results in the context of real-world applicability.

## **Week 5: August 21 - August 27**

**Task 8: Result Analysis and Final Report Drafting**

**Status: Remaining**

 **August 21 - August 23: Result Analysis**

* Conduct a thorough analysis of the model evaluation results.
* Identify key insights and findings from the analysis.

 **August 24 - August 27: Draft the Final Report**

* Compile the project findings into a detailed draft report.
* Include the methodology, results, conclusions, and suggestions for future work.

## **Week 6: August 28 - September 3**

**Task 9: Result Analysis and Final Report Drafting**

**Status: Remaining**

 **August 28 - August 31: Final Report Review**

* Conduct a thorough review of the draft report.
* Make necessary revisions based on feedback and refine the report for clarity and completeness.

 **September 1st – September 3rd: Final Review and Submission**

* Finalize the report and prepare for submission.
* Submit the final report.

# **Bibliography**

World Health Organization. "Cardiovascular Diseases (CVDs)." Retrieved from WHO

LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." Nature, 521(7553), 436-444.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., ... & van Ginneken, B. (2017). "A survey on deep learning in medical image analysis." Medical image analysis, 42, 60-88.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." arXiv preprint arXiv:1711.05225.

Chollet, F. (2017). "Xception: Deep Learning with Depthwise Separable Convolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1251-1258.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning." Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 4278-4284.

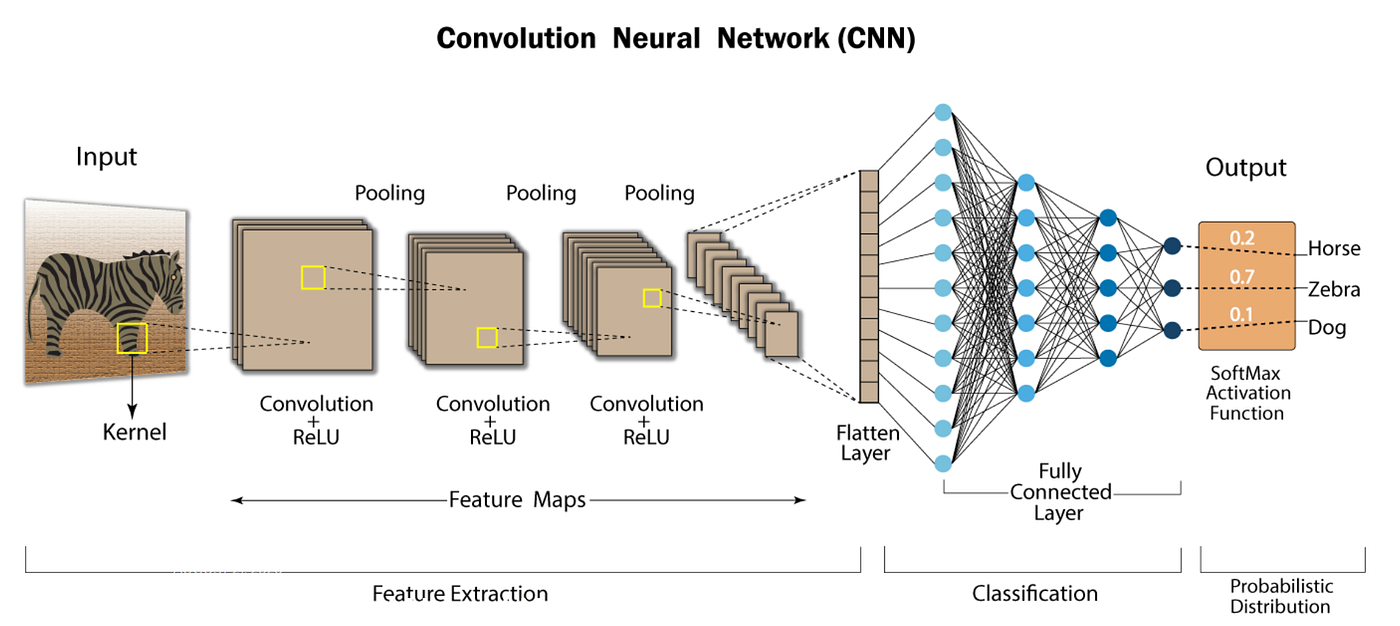
Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). "Dermatologist-level classification of skin cancer with deep neural networks." Nature, 542(7639), 115-118.

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning." Cell, 172(5), 1122-1131.

Lundberg, S. M., & Lee, S. I. (2017). "A Unified Approach to Interpreting Model Predictions." Advances in Neural Information Processing Systems, 4765-4774.

# **Appendices**

## **Appendix 1: Detailed CNN Architectures Explored**



## **Appendix 2: Data Preprocessing Techniques**

**Handling Missing Values**

* Method: Imputation using mean/median for continuous variables, mode for categorical variables
* Rationale: Ensures complete dataset for training without introducing bias

**Data Normalization**

* Method: Min-max scaling to range [0, 1] or standardization (z-score normalization)
* Rationale: Improves model convergence and training stability

**Categorical Variable Encoding**

* Method: One-hot encoding for nominal variables, label encoding for ordinal variables
* Rationale: Makes categorical data suitable for input to CNNs

## **Appendix 3: Evaluation Metrics Definitions**

**Accuracy**

* Formula: (True Positives + True Negatives) / Total Samples
* Interpretation: Proportion of correctly predicted samples

**Precision**

* Formula: True Positives / (True Positives + False Positives)
* Interpretation: Proportion of positive predictions that are actually positive

**Recall (Sensitivity)**

* Formula: True Positives / (True Positives + False Negatives)
* Interpretation: Proportion of actual positives that are correctly predicted

**F1-Score**

* Formula: 2 \* (Precision \* Recall) / (Precision + Recall)
* Interpretation: Harmonic mean of precision and recall, balances both metrics