Literature Review

I. Introduction

With an emphasis on early chronic disease detection and personalized medicine, predictive analytics in healthcare has emerged as a major field for study and application. The purpose of this literature review is to investigate foundational publications, ongoing projects, and approaches that are pertinent to the proposed healthcare analytics project.

II. Previous Work

Sigfried Gold, Noemie Elhadad, and James J Cimino [1] presented an innovative approach for extracting structured medication event information from discharge summaries. The method involves a sequential process starting with the definition of the targeted medication events, followed by the construction of a parser and the generation of parsing rules. To validate the effectiveness of the system, a test dataset is curated, and parsing rules are applied, with the intention of seeking input from physicians for final decision-making on medications. Testing on the created dataset yields defined scores. This method offers a systematic way to extract pertinent medication information, ensuring a comprehensive understanding. However, potential challenges may arise in cases where parsing rules may need frequent updates to adapt to variations in discharge summaries. The involvement of physicians in decision-making adds a valuable clinical perspective, but it may also introduce subjectivity into the interpretation of results.

Miotto et al. introduced DeepPatient [2], a system designed to harness raw patient data from Electronic Health Records (EHR), including information on medications, diagnoses, procedures, and lab tests. The system generates patient representations using unsupervised deep feature learning algorithms, opening the door to advanced clinical activities including medication targeting, clinical trial recruitment, and patient likeness detection. A notable strength of this research lies in its ability to leverage complex EHR data to create meaningful patient representations, allowing for sophisticated clinical applications. However, the application of deep learning algorithms might result in models that are challenging to interpret, raising concerns about data privacy and computational complexity. Despite these challenges, the system's versatility and potential for personalized clinical interventions make it a promising tool for enhancing healthcare outcomes.

With a focus on feature selection and classification methods, Jain D., Singh V.[3] offer an overview of the critical area of chronic illness prediction in healthcare. The significance of feature selection is emphasised in terms of enhancing the precision of classification algorithms through the identification and removal of superfluous features from illness datasets. The study explores traditional approaches such as filter, wrapper, and embedded methods and discusses hybrid approaches that bring together the best features of several approaches. The need of early detection for chronic conditions like diabetes, heart disease, and cancer is emphasised. The paper explores many feature selection algorithms, including their features, benefits, and drawbacks. Benefits include lower dimensionality and improved accuracy; drawbacks include possible biases and difficulties with big datasets.

Ghosh P., Azam S., Karim A., Jonkman M., Hasan M. D. Z[4], address the global challenge of cardiovascular disease-related deaths by presenting an intelligent diagnostic framework for heart disease prediction, utilizing the Cleveland Heart disease dataset. The study uses machine learning methods, namely Random Forest (RF), K-Nearest Neighbour (KNN), and Decision Tree (DT), considering three distinct feature sets: the whole collection, ten characteristics chosen using Pearson's Correlation, and six features chosen using the Relief algorithm. Among the assessment measures are sensitivity, accuracy, and precision. With Relief-selected characteristics, the Random Forest classifier produced the best results, with an accuracy of 98.36%, which was further increased to 99.337% using 5-fold Cross Validation. Although the study shows encouraging results for effective diagnosis of heart disease, future applications may need to consider potential constraints such dataset-specific biases and the requirement for external validation.

III. References:

1. Sigfried Gold, Noemie Elhadad and James J Cimino, “Extracting Structured Medication Event Information from Discharge Summaries”, *JAMIA symposium proceedings*, 2008.
2. R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, “Deep patient: An unsupervised representation to predict the future of patients from the electronic health records,” Sci. Rep., 2016.
3. Jain D., Singh V., “Feature selection and classification systems for chronic disease prediction: a review”. *Egyptian Informatics Journal,* 2018.
4. Ghosh P., Azam S., Karim A., Jonkman M., Hasan M. D. Z. “Use of efficient machine learning techniques in the identification of patients with heart diseases”, Proceedings of the 2021 the 5th International Conference on Information System and Data Mining, 2021.