COMP1682 – Final Year project
Project Report - Feedback sheet

Student Name and ID: Rifath Ahmed, 001187671
Second Marker Name: IK SOO LIM

Overall assessment

Title: Cardiovascular Disease Prediction with Advanced Machine Learning Techniques: A Comprehensive Analysis

This project demonstrates great promise in utilising recent machine learning (ML) algorithms for the prediction of cardiovascular disease (CVD). The study employs three distinct data sets for robust training and testing. The thorough literature review and self-contained description of the utilised ML algorithms are noteworthy. However, there is room for improvement.
The comparative study is somewhat limited, focusing on merely two ML algorithms. It would have been beneficial to conduct a more comprehensive comparative study, incorporating a wider range of ML algorithms whose source code or library code are readily available. This would have enabled a more detailed evaluation, as exemplified in several cited references.
The interpretation of the results could be more robust. The difference in performance between the two ML algorithms is, in fact, negligible. However, the student has attempted to imply that one is superior to the other based on this marginal difference.
It is worth noting that a previous study has employed the same data set and algorithm. A comparison between the results of this dissertation and the previous study would have been invaluable.

<<this study investigates the application of machine learning (ML) approaches to increase the accuracy of cardiovascular disease (CVD) prediction and diagnostics. The paper utilised two complex machine learning models, 1D-CNN and TabNet, to handle class imbalances.>>

Why focus on only two ML algorithms? A more comprehensive comparative study involving a broader range of ML algorithms would have been beneficial.

<<Traditional machine learning models, such as decision trees (DT), random forests (RF), and support vector machines (SVM), were utilised in the early stages of the cardiovascular disease prediction process… The development of deep learning, which provides the capability to learn hierarchical representations of data automatically, was a big step forward in machine learning.>>

It is essential to acknowledge that more recent algorithms do not necessarily yield improved performance compared to traditional ones. Therefore, a thorough comparative study would be advantageous, incorporating both recent and traditional ML algorithms and highlighting the extent of differences in their performance, as demonstrated in recent references (Bárbara Martins, 2021; PRONAB GHOSH, 2021; AZAM MEHMOOD QADRI, 2023; Rout, 2023).

Detailed descriptions of the two ML algorithms are provided.

<<After thorough analysis, it is evident that the 80-20 split ratio consistently proves to be the most effective across all datasets and models.>>

Fine.

<<From the results presented, TabNet generally demonstrates a slight advantage over 1D-CNN in precision and F1 scores across all datasets, indicating its superior ability to minimise false positives and maintain a balanced approach between precision and recall.>>

A more robust conclusion would be that the two ML algorithms exhibit similar performance, differing by a mere 0.01 in the performance metric.

<<1D-CNN, however, exhibits a strong recall performance, which is particularly useful for ensuring actual cases of heart disease are not missed.>>

This conclusion may not be convincing. Based on the results from the three data sets – (1D-CNN, TabNet) = (0.83,0.83), (0.85,0.83), and (0.98,0.99) – 1D-CNN outperforms TabNet in only one data set by a marginal 0.02. A more robust conclusion would be that they exhibit similar performance.

<< (Raniya R. Sarra, 2022) … Cleveland … 1D-CNN…>>

One of the references employs the same Cleveland data and 1D-CNN, yet demonstrates differing performance compared to that reported in this dissertation – for instance, accuracy: 87% vs 84%, F1-score: 87% vs 82%. Surprisingly, there is no attempt to compare these results, despite this project utilising the same ML method and data set.

Comments on the viva/demonstration

Have you held a demonstration or viva for this project? YES /

When creating a poster, it's more effective to use concise language, such as bullet points, rather than lengthy sentences. Additionally, it would be beneficial to place greater emphasis on the key advantages or innovations of your work, highlighting how it improves upon existing relevant research.

Comments on the Standard of Presentation

Has the report met presentation criteria? YES

<<Alanazi, A., 2022. Using machine learning for healthcare challenges and opportunities. ScienceDirect, Volume 30.>>

Some references contain inaccurate information. For instance, ScienceDirect is not a journal name but the name of a portal/publisher. The correct name of the journal for this example

should be 'Informatics in Medicine Unlocked'.

<<34. Serkan Kiranyaz, O. A. O. A. T. I. M. G. D. J. I., 2021. 1D convolutional neural networks and applications: A survey. Mechanical Systems and Signal Processing, Volume 151.>>

Some of references do not have the required full information such as missing issue and page numbers.