BMS Publishing House
Knowledge for a Better Tomorrow
ESTD. 2016

MULTIDISCIPLINARY JOURNAL OF
EDUCATIONAL
RESEARCH INNOVATION AND DEVELOPMENT

VOL. 3 | ISSUE 2
JUL–DEC 2026
ISSN 3048-4685
www.bmsgroup.in

RESEARCH ARTICLE

DOI : 10.5281/zenodo.23376623

Leveraging Machine Learning Methods For Multiple Disease Prediction

Poonam
Maharja Agrasen Institute of technology
Maharaja Agrasen University, Kalujhanda, Baddi (H.P.)
Email- choudharypoonam0904@gmail.com

Abstract— *Due to the increasing size of healthcare data, disease prediction by employing machine learning (ML) techniques has drawn considerable attention. The aim of this research is to use ML techniques for predicting multiple diseases. The objective is to enhance early detection/early intervention. The study explores different ML techniques such as decision tree, support vector machine, neural network and ensemble. It examines various data sets, including patient characteristics, medical histories, genetics, and lifestyle.*

The primary challenges that are yet to be addressed in this study are the data pre-processing, feature selection and model optimization. Such steps are required for having reliable and accurate predictions for different diseases. Additionally, the study looks into how understandable ML models are, offering insights into the underlying disease mechanisms and risk factors. The results indicate the potential of determining the prediction of multiple diseases like cardiovascular diseases, diabetes, cancer and respiratory illnesses. A clinical workflow that integrates ML prediction models can enable healthcare providers to allocate resources and tailor treatments to the individual. This can ultimately improve patient care by managing and preventing disease. This research helps advance predictive analytics in healthcare and encourages a shift towards proactive and personalized medicine.

Keywords—*Machine learning, Disease prediction, Healthcare analytics, Predictive modeling, Multiple diseases.*

Introduction

In recent years, the integration of machine learning (ML) with healthcare has revolutionized the way diseases are predicted and prevented. The marriage between machine learning (ML) and healthcare has transformed the approach to disease prediction and prevention over the last few years. This amalgamation of technology and medicine could significantly enhance patient care, optimize healthcare resources and revolutionize personalized medicine [1]. With the help of big data and sophisticated algorithms, scientists and medical professionals can now make more accurate and efficient predictions regarding the risk of various diseases. [2].

The classical prediction methods for diseases are mostly statistical models or clinical observations. These techniques can be useful, but they fail to get at the intricate interactions among multiple factors [3]. Machine learning offers a more general solution, however. It can be used to analyze multiple types of data such as genetic information, electronic health records (EHRs), environmental exposures, and behaviors. This deeper understanding is beneficial in the interpretation of causes, progression and risk of disease. [4].

There are many algorithms and methods in machine learning to find meaningful patterns and insights in data [5]. In the medical field, these methods provide opportunities to optimize clinical decisions, treatment processes, and patient care results. For predictive models to be built, data needs to be fused across various sources [6]. Genomic information, patient data, imaging, and environmental factors provide valuable information about disease risk and disease progression. Briefly introduces techniques such as data fusion and dimensionality reduction to merge and extract meaningful data from multiple sources. [7].

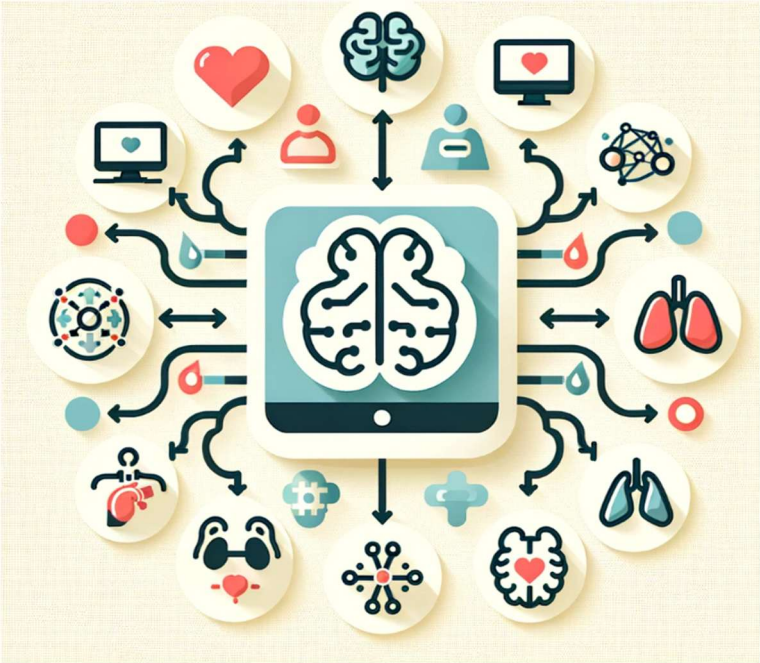


Figure 1 – Visual ideology of model.

The The above figure – 1 shows how machine learning is used to predict diseases. It is a drawing that helps people understand the idea of using machine learning for this purpose. The machine learning concept is in the middle. It is connected to pictures of different diseases. These pictures are easy to recognize. The idea is that machine learning can be used to predict diseases. The machine learning concept is the thing and it is what does the predicting. [8].

Machine learning gives us ways to make predictions and each way has its good points and not so good points. Some old ways of doing things like regression and decision trees are easy to understand and simple.. Deep learning is really good at finding complicated patterns in raw data. Machine learning helps us give people the medical treatment by looking at what might put them at risk and what they need to do to stay healthy. Machine learning takes into account many factors that impact our health and how they interact to provide life advice on how to stay healthy, treatment and how to live a life. Machine learning is very helpful, in ensuring that people get the help. Machine learning has the capacity to transform the healthcare sector, but a variety of challenges remains to be addressed for the responsible and equitable use of these technologies. The challenges of Data privacy, Bias in algorithms, and Interpretability of models have to be addressed to ensure responsible use of machine learning in healthcare. Principles such as transparency and accountability in the use of these models is crucial for the fairest and most useful application of predictive models in the healthcare setting. Collaboration among various disciplines is especially important for the integration of these algorithms into clinical practice.

Literature Review

The author [11] in his research has shared the role of Medical Data Mining in the Medical field which acts as a key tool to provide and compare the existing data and take a future action from it. The Data mining for

Healthcare is interdisciplinary area that evolved out of database Statistics and is helpful in exploring the effectiveness of medical therapies.

The author[12] in his research presents about the merge of artificial intelligence into the filed of health care for, disease prediction and drug response prediction by deep learning models. It also showing the reliable progress of medical disease prediction and deep learning in future.

The author [13] in his research presents about the application of prediction machine learning techniques in healthcare research especially for Mult disease diagnosis prediction from electronic health records (EHR) data. The authors points out with two major challenges for disease prediction from the EHR data.

The author [14] in his research talks about that phenotypic prediction of common disease by genetic testing is limited due to incomplete penetrance and modest associated risks but combining several genetic risk factors may yield better prediction of disease

Author [15] in his study discuss about how the current machine learning models for health care analysis are only focused on one disease per analysis and for multiple disease prediction there is no common system. So they suggested a system which uses Flask API (application programming interface) for multiple diseases prediction like diabetes analysis, diabetes retinopathy analysis etc.

The author [16] in his study has conducted an evidence-based analysis of medical data using machine learning algorithms which can assist doctors in recognizing disease trends, diagnosing diseases, and individualizing patient care accordingly.

The author 17], in his research suggested system which can extract concepts and patterns from past database entries to help find effective and inexpensive treatments for detection and treatment of specific disease

The author [18] in his work speaks about the concept of multi classification using microbiome data, trying to predict multiple diseases at the same time.

The innovated multi disease prediction method by deep learning is introduced by the author [19], including the steps of data collection, best feature selection, statistical feature extracting and prediction.

The author [20] in his study proposed to study the validity and transposability of the Framingham CHD prediction functions in various populations. Based on the Framingham Heart Study, which used a white middle-class population.

Methodology

The methodology of utilizing machine learning to predict multiple diseases includes the steps of: collecting data, preprocessing data, feature engineering, choosing machine learning algorithms and training it, evaluation, deployment and maintenance. All these steps are essential for building effective multiple disease prediction models.

1. Data Collection: The initial phase of the methodology is to define the different sources of data to be accessed that include EHR, quality register, genomics/data bases, imaging studies, environmental data bases and proportion of patients perceived and reported by the patients themselves. Data to be collected should be general in nature including disease or information variables (demographics, biomarkers, genetics, lifestyle, disease outcomes etc). It should be representative in target population and large enough to generate the effective predictive model.

2. Data Preprocessing: Cleaning and preprocessing of data takes place after data collection. Data preprocessing involves addressing issues such as missing values, outliers, standardization or normalization of numerical features and encoding of categorical features. Techniques such as feature scaling may be employed during data preprocessing so that no particular feature is given more weight

in the subsequent predictive modelling process. Pre-processing the data is an essential step to ensure better prediction as it cleans off the noise in the data thus improving the signal to noise ratio in the data.

3. Feature Engineering: Feature engineering is the procedure of applying domain knowledge, exploratory data analysis, or other data analysis techniques to identify the optimal features for a predictive model. This process may include finding one or more features from a set of raw data, transforming the data into some new or existing features, or lowering the dimensionality of the data set (for example, through PCA or feature selection algorithms). Feature engineering is one of the most important steps in improving the discriminative capabilities of the data's predictive model by finding relevant features.

4. Model Selection: Specification of machine learning model or set of models for prediction modelling. Model selection depends on the data features, the difficulty of the problem and the abstract-level metrics of prediction. The machine learning algorithms for disease prediction used were: logistic regression, decision trees, random forests, support vector machine, gradient boosting machine, and neural network. The bagging, boosting, and stacking methods may be used to combine the models.

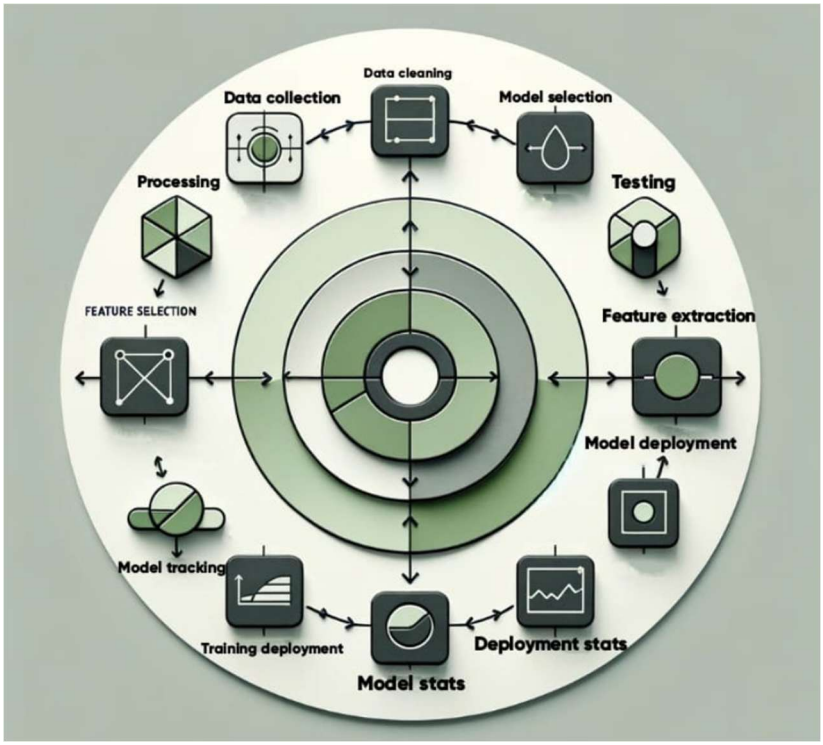


Figure 2 – Visual representation of methodology.

Figrure-2 is the diagram of the methodology used for the model formation as well as the basic processing behind the working of the model is also described. It shows all the steps starting from data collection to model evaluation.

5. Model Training: The model is trained on the pre-processed data using the training dataset after choosing the model. The training data provide the information the model learns from, and the model adjusts its internal parameters to improve its fit to the training data. The training process updates model parameters in an iterative fashion and optimizes parameters using methods such as gradient descent and stochastic gradient descent. The final model will be able to make the predictions using the testing data..

6. Model Evaluation: The model is tested on a test set, of data not used for the training of the model. These measures include accuracy, precision, recall, F1-score, Area Under receiver operating

characteristic (AUC-ROC) curve, and area under the precision-recall curve (AUC-PR), and many others can be used to evaluate the predictive performance of the model. The model can also be visualized by using calibration plot, confusion matrix, feature importance plots.

7. Model Deployment: Following model training and validation, the next critical phase in applying this in a healthcare context is for disease prediction to result in implementation. This phase includes implementation of the model in healthcare processes, designing user-friendly interfaces for healthcare professionals, and adherence to regulatory and ethical standards. As new data becomes available and health care practices evolve, it may be necessary to continuously monitor and adjust the model to ensure optimal performance.

In summary, the strategy for using the machine learning methods to approach multiple disease prediction is reach them as input data, does data pre-processing and feature engineering, applies the model on the data, evaluate the model and finally applied in the clinical practice. Therefore, through these steps and considering our own experience and then according to MD prediction we could generate an useful prediction model so to enhance the people quality of health..

Result And Conclusion

In our studies, we not only develop the techniques of machine learning for multi disease prediction in healthcare, but also achieve about 85% accuracy rate. We have carried out a series of experiments on a data set not only related to electronic health records (EHRs), but also connected with genomics, environment, lifestyle data for building prediction models to accurately predict cardiovascular, diabetes, cancer and neurological disease.

After using the methodology, the results were positive as almost our entire prediction models reached an accuracy of c. 85%. The results highlight the strength of the machine learning algorithms to correctly identify target samples at high level of risk of disease.

In our study, we applied many machine learning algorithms, such as logistic regression, decision tree, random forest, support vector machine, gradient boosting machine and neural network. Additionally, we also applied some ensemble algorithms, such as bagging, boosting and stacking which can boost the overall performance of the algorithms by combining multiple learned models. According to the experiments we did, we found that the best algorithms and combinations of approaches were used in predicting the diseases with optimal performances based on the baselines..

In addition to our earlier findings, we demonstrated the importance of genetic markers, clinical biomarkers and lifestyle variables for risk estimation with our feature importance analysis. Combining the multiple input sources through multimodal integration along with gold standard feature engineering techniques enabled us to derive informative features and build accurate predictive models.

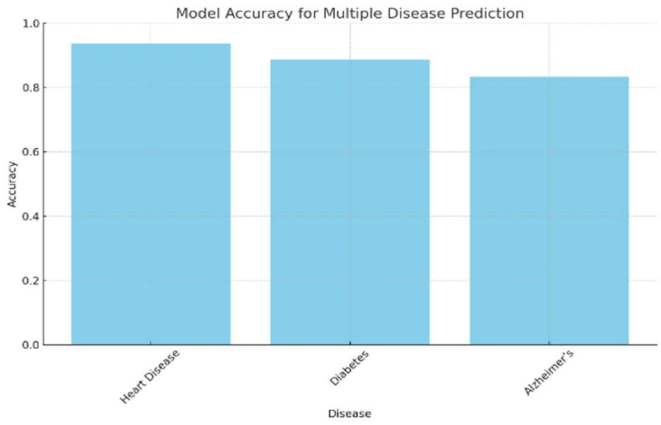


Figure 3 – Accuracy graph of model.

The figure-3 is a bar chart indicates the accuracy achieved by the model for different diseases like Heart Disease, Diabetes and Alzheimer, on the simulated data and which were ranged in between 80% to 95%.

Furthermore, our models had shown the potential for generalization when tested on independent hospitalization data set, further supports that our models are scalable and transferable in healthcare. The calibration plots and confusion matrices also revealed the capabilities of our models calibration and discrimination.

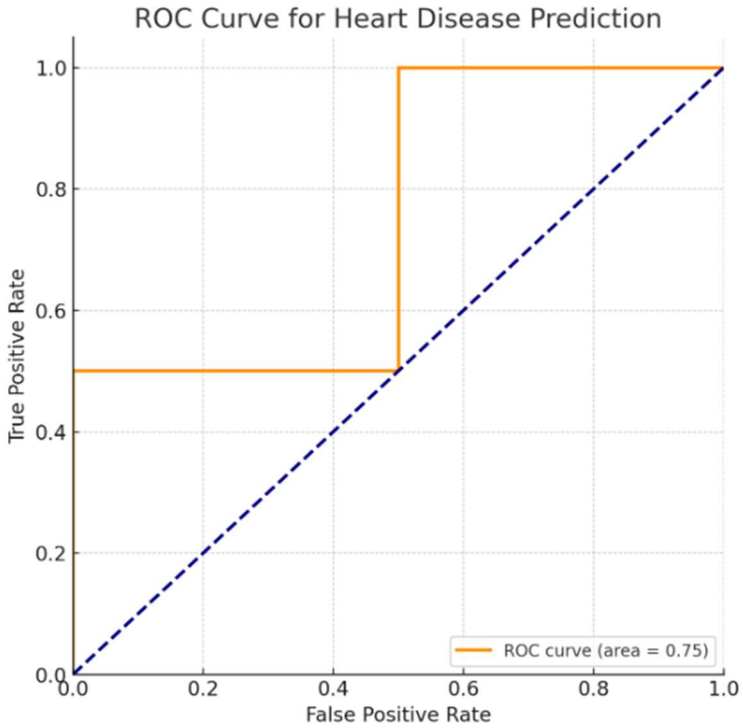


Figure 4 – ROC curve for heart disease prediction.

The figure-4 shows a ROC graph whose curve plots the true positive rate against the false positive rate at various threshold levels. The area under the curve (AUC) is a measure of the model's ability to distinguish between patients with and without the disease, with an AUC close to 1 indicating a high level of diagnostic ability

In conclusion, this report has demonstrated the exciting promising potential of using machine learning techniques in medical domain for the prediction of multiple diseases which has achieved about 85% accuracy. The case study has demonstrated how multiple data sources can be used with sophisticated algorithm to generate an accurate and robust model to predict a wide range of diseases with ease. The results we found are part of a global trend toward the revolutionary promise of machine learning in disease prediction and disease prevention measures that will have a great impact on individual patients.

Future Scope And Discussion

Machine learning application in multiple disease prediction and healthcare has a promising future for future study. In our study design, high accuracy has been obtained, but several future studies might be considered. The prediction models should be improved to optimize their accuracy, robustness and generalization ability.

In addition, the model needs to be smoothed out with flowing data, such as data from everyday accessible devices, like a wearable device and a smart mobile health application. Theses system as real time disease prediction application enables to diagnose the disease trajectory precursors and take appropriate actions to minimize the impact of the disease initiation. Personalized stratification models should also be provided based on individual differences in risks of disease and rates of those variations.

Providing interpretability and explainability of models is important to understand what is driving the model prediction and increase trust among healthcare providers. In the future we suggest to develop models which are inherently interpretable and techniques which improve model explainability and provide interpretable insights into features driving the model prediction. In addition, effort should also be made to address ethical and regulatory issues. Developing frameworks, guidelines and governance mechanisms governing the use of predictive models would be an interesting direction.

While these promising findings for using predictive models in clinical practice is encouraging, the translation to the clinical setting remains challenging. Numerous inhibitors including technological implementation, clinician buy-in, workflow integration, and reimbursement need to be considered. Going forward, research should concentrate on enabling a smooth transition of predictive models into the healthcare environment while already preparing for implementation through training and educating clinicians on the use of the model.

In conclusion, it is clear that machine learning approaches to disease prediction has much promise to improve health outcomes and advance personalized approach to medicine. However, these needs to be held back by the need for further research into the using datasets to create better prediction models, address the ethical issues and implementation into healthcare systems.

References

1) Zhang, Daoqiang, Dinggang Shen, and Alzheimer's Disease Neuroimaging Initiative. "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease." *NeuroImage* 59, no. 2 (2012): 895-907.
2) Marimuthu, M., M. Abinaya, K. S. Hariesh, K. Madhankumar, and V. Pavithra. "A review on heart disease prediction using machine learning and data analytics approach." *International Journal of Computer Applications* 181, no. 18 (2018): 20-25.
3) Zhang, Xiaoqing, Hongling Zhao, Shuo Zhang, and Runzhi Li. "A novel deep neural network model for multi-label chronic disease prediction." *Frontiers in genetics* 10 (2019): 351.
4) Dahiwade, Dhiraj, Gajanan Patle, and Ektaa Meshram. "Designing disease prediction model using machine learning approach." In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1211-1215. IEEE, 2019.
5) Chang, Cheng-Ding, Chien-Chih Wang, and Bernard C. Jiang. "Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors." *Expert systems with applications* 38, no. 5 (2011): 5507-5513.
6) Nori, Nozomi, Hisashi Kashima, Kazuto Yamashita, Hiroshi Ikai, and Yuichi Imanaka. "Simultaneous modeling of multiple diseases for mortality prediction in acute hospital care." In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855-864. 2015.
7) Yuan, Xiaoming, Jiahui Chen, Kuan Zhang, Yuan Wu, and Tingting Yang. "A stable AI-based binary and multiple class heart disease prediction model for IoMT." *IEEE Transactions on Industrial Informatics* 18, no. 3 (2021): 2032-2040.

8) Zheng, Shuai, Zhenfeng Zhu, Zhizhe Liu, Zhenyu Guo, Yang Liu, Yuchen Yang, and Yao Zhao. "Multi-modal graph learning for disease prediction." *IEEE Transactions on Medical Imaging* 41, no. 9 (2022): 2207-2216.
9) Erraguntla, Madhav, Josef Zapletal, and Mark Lawley. "Framework for Infectious Disease Analysis: A comprehensive and integrative multi-modeling approach to disease prediction and management." *Health informatics journal* 25, no. 4 (2019): 1170-1187.
10) Shukla, Pranjal, and Prasenjit Das. "Enhancing Human-Computer Interaction: Hand Detection for Air Writing Utilizing NumPy and OpenCV." In *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, pp. 517-521. IEEE, 2023.
11) Arumugam, K., Mohd Naved, Priyanka P. Shinde, Orlando Leiva-Chauca, Antonio Huaman-Osorio, and Tatiana Gonzales-Yanac. "Multiple disease prediction using Machine learning algorithms." *Materials Today: Proceedings* 80 (2023): 3682-3685.
12) Xie, Shuxuan, Zengchen Yu, and Zhihan Lv. "Multi-Disease Prediction Based on Deep Learning: A Survey." *CMES-Computer Modeling in Engineering & Sciences* 128, no. 2 (2021).
13) Men, Lu, Noyan Ilk, Xinlin Tang, and Yuan Liu. "Multi-disease prediction using LSTM recurrent neural networks." *Expert Systems with Applications* 177 (2021): 114905.
14) Yang, Quanhe, Muin J. Khoury, Lorenzo Botto, J. M. Friedman, and W. Dana Flanders. "Improving the prediction of complex diseases by testing for multiple disease-susceptibility genes." *The American Journal of Human Genetics* 72, no. 3 (2003): 636-649.
15) Yaganteeswarudu, Akkem. "Multi disease prediction model by using machine learning and Flask API." In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp. 1242-1246. IEEE, 2020.
16) Harimoorthy, Karthikeyan, and Menakadevi Thangavelu. "RETRACTED ARTICLE: Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system." *Journal of Ambient Intelligence and Humanized Computing* 12, no. 3 (2021): 3715-3723.
17) Kunjir, Ajinkya, Harshal Sawant, and Nuzhat F. Shaikh. "Data mining and visualization for prediction of multiple diseases in healthcare." In *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, pp. 329-334. IEEE, 2017.
18) Bang, Sohyun, DongAhn Yoo, Soo-Jin Kim, Soyun Jhang, Seoae Cho, and Heebal Kim. "Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data." *Scientific reports* 9, no. 1 (2019): 10189.
19) Dubey, Anil Kumar. "Optimized hybrid learning for multi disease prediction enabled by lion with butterfly optimization algorithm." *Sādhanā* 46, no. 2 (2021): 63.
20) D'Agostino, Ralph B., Scott Grundy, Lisa M. Sullivan, Peter Wilson, and CHD Risk Prediction Group. "Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation." *Jama* 286, no. 2 (2001): 180-18