A Machine Learning Approach towards Detecting Dementia based on its Modifiable Risk Factors

Reem Bin-Hezam[1]

Information Systems Department
College of Computer & Information Sciences
Princess Nourah bint Abdulrahman University
Riyadh, Saudi Arabia

Tomas E. Ward[2]

Insight Centre for Data Analytics
Dublin City University
Glasnevin, Dublin 9, Ireland

for the Alzheimer's Disease Neuroimaging Initiative*

Abstract—**Dementia is considered one of the greatest global health and social care challenges in the 21st century. Fortunately, dementia can be delayed or possibly prevented by changes in lifestyle as dictated through known modifiable risk factors. These risk factors include low education, hypertension, obesity, hearing loss, depression, diabetes, physical inactivity, smoking, and social isolation. Other risk factors are non-modifiable and include aging and genetics. The main goal of this study is to demonstrate how machine learning methods can help predict dementia based on an individual's modifiable risk factors profile. We use publicly available datasets for training algorithms to predict participant's cognitive state diagnosis, as cognitive normal or mild cognitive impairment or dementia. Several approaches were implemented using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) longitudinal study. The best classification results were obtained using both the Lancet and the Libra risk factor lists via longitudinal datasets, which outperformed cross-sectional baseline datasets. Moreover, using only data of the most recent visits provided even better results than using the complete longitudinal set. A binary classification (dementia vs. non-dementia) yielded approximately 92% accuracy, while the full multi-class prediction performance yielded to a 77% accuracy using logistic regression, followed by random forest with 92% and 70% respectively. The results demonstrate the utility of machine learning in the prediction of cognitive impairment based on modifiable risk factors and may encourage interventions to reduce the prevalence or severity of the condition in large populations.**

Keywords—*Machine learning; classification; data mining; data preparation; dementia; modifiable risk factors*

I. INTRODUCTION

Dementia presents enormous global health and social challenges. Currently, there are around 47 million people with dementia worldwide, and that number is expected to triple by 2050. Dementia occurs mainly in people older than 65 years [1]. The aging population worldwide is almost certainly part of the reason behind this increase, especially in low- and middle-income countries.

Dementia is a collection of symptoms of cognitive defects, which could be delayed or possibly prevented by eliminating certain modifiable risk factors associated with the condition. However, few researches used machine learning approaches to detect dementia based on its modifiable risk factors, while most of the previous researches used machine learning to

detect dementia based on imaging data or non-modifiable factors such as genetics. Although these methods are useful in diagnosing dementia, they may not be as much useful in term of delaying or preventing dementia as there is nothing that could be modified.

This study aims to use a machine learning (ML) approach to classify the cognitive state and detect dementia based only on the modifiable risk factors. The main research objective is to determine to what extent it is possible to predict dementia based on an individual's modifiable risk factors profile.

The analysis of this research is applied to data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) longitudinal study, using modifiable risk factors lists that have been already defined by the *Lancet* commission and the *Libra* index. As far as known, no previous work has explored *Lancet,* and *Libra* lists of modifiable risk factors on the ADNI dataset using a machine learning approach.

Moreover, being able to accurately detect dementia based only on its modifiable risk factors would make it possible not only to predict dementia but also to target its risk factors. This will be useful in term of trying to delay or prevent the disease by eliminating these factors as possible.

The remaining of this paper is structured as follows. Section II provides background on the domain and some related work. The methodology applied in this research is described in Section III. Moreover, the experiment and results are provided in Section IV. Finally, the conclusion of the research and its future work is provided in Section V.

II. BACKGROUND

Dementia is described as a collection of symptoms related to cognitive deficits and is not considered one single disease. In the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [2], dementia is listed under Major Neurocognitive Disorder (NCD), and is defined by the following:

- There is evidence of a substantial cognitive decline in one or more cognitive domains.

- The cognitive deficits interfere with independence in everyday activities, are not exclusively in the context of a delirium, and are not mainly attributable to another mental disorder.

Risk factors for dementia can be either modifiable or non-modifiable [3]. Fortunately, dementia could be delayed or possibly prevented by eliminating some of its modifiable risk factors [4].

A. Dementia Risk Factors

The *Lancet* Commission study found that around 35% of dementia risk factors are potentially modifiable [1]. These risk factors include less education, hypertension, obesity, hearing loss, depression, diabetes, physical inactivity, smoking, and social isolation. Although the impact of these factors varies at different life stages, eliminating them at any stage would be beneficial. Moreover, studies recommend active treatment and intervention of modifiable dementia risk factors, which would potentially delay or prevent 30% of dementia cases [1], [4].

On the other hand, completely eliminating the apolipoprotein E (APOE) ε4 allele, which is considered the major genetic risk factor of dementia, could reduce its incidence by 7% [1]. However, this and all other genetic factors are considered to be non-modifiable. Besides genetics, other non-modifiable risk factors include age and gender.

A common method to calculate dementia risk based on its risk factors is by using the Lifestyle for Brain Health (LIBRA) index [5], [3], [6], which is calculated by the Innovative Midlife Intervention for Dementia Deterrence (In-MINDD) project [7].

Table I listed the modifiable risk factors defined by both *Lancet* commission and *Libra* index.

The National Academy of Medicine committee [8] also identified cognitive training, blood pressure management for people with hypertension, and increased physical activities as three main classes of dementia intervention.

Alzheimer's disease (AD) is the most common type of dementia. The next most common type is vascular dementia (VaD), followed by dementia with Lewy bodies. Frontotemporal degeneration and dementia associated with brain injury, infections, and alcohol abuse are less common types of dementia [1].

TABLE. I. DEMENTIA RISK FACTORS AND DIAGNOSIS ATTRIBUTES

#	Risk Factor	Lancet List	Libra List
1	Low Education	●	●
2	Hypertension	●	●
3	Obesity	●	●
4	Smoking	●	●
5	Depression	●	●
6	Diabetes	●	●
7	Physical inactivity	●	●
8	Hearing loss	●	
9	Social isolation	●	
10	Cognitive inactivity		●
11	Chronic Heart disease		●
12	Alcohol use		●
13	Chronic Kidney disease		●

Tariq and Barber [9] suggested dementia prevention by targeting vascular modifiable risk factors, as these two types are often co-existing in the brain and share some common modifiable risk factors.

B. Current Approaches used in Detecting Dementia Risk Factors

Many studies have aimed to predict an early diagnosis of dementia through magnetic resonance imaging (MRI) and genetic variables [10]. However, these measurements are expensive and not always available.

Most of the research that has used machine learning applied classification methods from MRI data to classify or predict a diagnosis of different cognitive diseases and states [11], [12], [13].

On the other hand, only a few studies have used machine learning techniques to determine risk factors associated with dementia or one of its major causes (i.e., Alzheimer's disease) [11]. Some of the studies combined modifiable and non-modifiable risk factors in order to reach a higher level of accuracy.

Most of the available research used large cohort studies and a population-based perspective to determine associated risk factors [14], while some used statistical analysis to provide a ranked risk-factor index [3], [6].

Two main studies used machine learning techniques to detect dementia's risk factors and predict dementia risk accordingly [15], [16]. Both studies applied their analysis to one longitudinal cohort study with a relatively small size (i.e., 840 and 746 subjects respectively).

O'Donoghue, et al. [15] applied a non-linear dementia survival prediction model with a multilayer perceptron (MLP), which is an artificial neural network (ANN), and used both modifiable and non-modifiable risk factors defined in the In-MINDD project [5]. They also examined the hidden layers to extract different clusters of risk factors and explore different interactions between them. Due to a class imbalance of the MAAS dataset, their models were able to predict survival better than predicting dementia. Their models overall accuracy ranges between 53.57% and 70.24%.

Joshi, et al. [16] tried different attribute-evaluation methods on the major risk factors of both Alzheimer's and Parkinson's diseases, which included both modifiable and non-modifiable risk factors. They used a relatively small dataset of fewer than 500 subjects from the ADRC and ISTAART studies [16]. Their attribute-evaluation methods included Chi-Squared, Gain Ratio, Info Gain, Relief F, and Symmetrical Uncertainty. They then applied several machine learning models, including Decision Tree, Random Forest (RF), and MLP to predict the patient's future status based on the defined risk factors. Their models did not detect dementia itself but instead classify subjects' diagnoses from three neurodegenerative diseases, which are AD, VaD, and Parkinson's.

Conversely, other studies aimed to predict dementia from neuroimaging data and in particular magnetic resonance imaging (MRI) or positron emission tomography (PET) scans

of the brain. Ding, et al. [17] were able to predict Alzheimer's disease around six years before its diagnosis using fluorine 18 fluorodeoxyglucose PET images of the brain. They achieved 82% specificity at 100% sensitivity using a deep learning algorithm. In another study, Casanova, et al. [12] used both MRI images and cognitive tests to detect Alzheimer's risk using regularized logistic regression.

Although prediction using MRI or PET scans or even genetics data can be very accurate, it is not practical in many countries to scale such an approach for population screening, and it does not present direct links with potentially modifiable factors that could be taken into account by an individual patient to delay dementia.

There have been no published studies to date investigating machine learning approaches with larger datasets to link modifiable risk factors to dementia and therefore providing suggestions for treatment and lifestyle change based on multiple population-based longitudinal studies. Modern machine learning methods over and above those used in the aforementioned studies focusing on modifiable risk factors and larger datasets should be explored to determine if they can produce better predictions and Insight.

Moreover, using possibly interpretable models in clinical research is essential for intervention development and for gaining an understanding of the relationships and interactions between symptoms or risk factors and diagnosis. Interpretability is difficult to achieve using black-box models such as neural networks, which contains hidden layers, although they might yield higher prediction accuracy. The easiest way to achieve interpretability is through interpretable models such as linear and logistic regressions, decision trees, and Naïve Bayes [18]. Consequently, this paper focuses on such methods with modifiable risk factors as input variables trained and tested on datasets significantly larger than those reported upon to date.

C. The Alzheimer's Disease Neuroimaging Initiative (ADNI) Study

Early prediction of dementia requires tracking changes in cognitive ability over time. The ideal study type which can support this tracking is one yielding longitudinal data points. In longitudinal studies, data are collected on one or more variables repeatedly, over time, in contrast with cross-sectional studies, in which data are collected on one or more variables at a single time point [19] [20].

The Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu) is a longitudinal study that was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. Its primary goal has been to test whether MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) [21].

The dataset consists of three longitudinal studies on around 1900 participants in total. ADNI enrolls participants who are between the age of 55 and 90 and are either normal healthy older adults used as controls (CN), people with either early or late MCI, and people with AD. The cognitive-state diagnoses (as well as dementia status) rating assessment of the participants was also provided.

The ADNI data set has been widely used in many research studies [11], [12], [13]. However, none of the published research that has used the dataset to date has attempted a machine learning approach to predict dementia based on established modifiable risk factors or even to explore the dataset for other possible dementia risk factors. Most of the research has instead focused on using MRI and PET scans or genetic data to predict Alzheimer's disease.

Most previous studies using the ADNI dataset and other longitudinal studies in the dementia field have used a complete case analysis (CCA) [22], and thus they considered only the cases with complete data and removed the missing values [11] [15] [23] [24]. Moreover, as per [25], if there is an overall worsening trend in health over time, missing data can be imputed from the same subject using their other available data.

While researches have shown the importance of preventing or delaying dementia, which might be achieved by targeting known modifiable risk factors, few studies have applied machine learning approaches to selecting dementia's risk factors and predicting dementia status. However, some studies combined both modifiable and non-modifiable risk factors.

More research and work in this area would improve the early prediction of dementia and recommend actions that would possibly prevent or at least delay its onset by targeting only the non-modifiable risk factors. Using an interpretable machine learning approach on the attribute selection and prediction would help to predict dementia based on its modifiable risk factors.

The main research contribution of this study is a demonstration of the utility of interpretable machine learning methods for the purposes of predicting future cognitive status for an individual based on modifiable risk factors that have been already defined by the *Lancet* commission and the *Libra* index.

III. METHODOLOGY

This research follows one of the most widely used process models for predictive data analytics, which is the Cross-Industry Standard Process for Data Mining (CRISP-DM) model adapted from [26] (see Fig. 1). The project lifecycle phases, as illustrated in the diagram, are business understanding, data understanding, data preparation, modeling, evaluation, deployment, and monitoring. All phases are going to be included in this project except deployment and monitoring, which are beyond the scope of this research. Domain understanding has already been established in the background (Section II).

A. Understanding the ADNI Dataset

The ADNI dataset is extensive, containing hundreds of tables with different categories from primary patients' demographics to highly complicated genes and imaging datasets; however, not all tables were useful for the scope of this research. Therefore, an initial investigation of the dataset

and its categories, subcategories, tables, fields, and their descriptions were needed. Fortunately, ADNI provides a data dictionary and an inventory that describe each table and its fields. The risk factors features are the independent variables while the diagnoses of the cognitive state are the independent variable, which might be one of three: cognitive normal (CN), mild cognitive impairment (MCI), and Dementia.

1) The modifiable risk factor attributes in ADNI: As the ADNI study dataset is extensive and consists of hundreds of tables and features under multiple categories, which may not be needed or useful for the aim of this research, the data dictionary, and the inventory were used to track only the necessary tables and features within them.

After reviewing the tables listed, the attributes related to dementia risk factors and diagnoses were selected. These attributes are listed in Table II. Attributes were selected from all ADNI cohorts except ADNI3 as the protocol of taking the medical history was different, and thus, not all features were available. A total of 1812 subjects were considered in the analysis.

2) Cross-Sectional vs. longitudinal data: As the dataset used in this research is longitudinal, another step was needed to understand the data through the study timeline. First, an understanding of how the data appear as cross-sectional, either at the baseline or at any single time point, was obtained. Then, a complete longitudinal view of the dataset was analyzed, including the differences between the main study parts (i.e., ADNI 1, Go, 2, and 3) and each visit's collected data.

B. Data Preparation

In this phase, the data were prepared for modeling by applying various data mining techniques to clean and to preprocess the data. This includes handling missing values, feature extractions, features transformation, and other tasks. Dealing with longitudinal data adds a complexity level to the preparation process because there could be various reasons and explanations for the data over time. A summary of the data preparation steps is shown in Fig. 2.

Fig. 1. CRISP-DM Model for the Project Phases (Adapted from [26]).

TABLE. II. MODIFIABLE RISK FACTORS AND DIAGNOSIS ATTRIBUTES

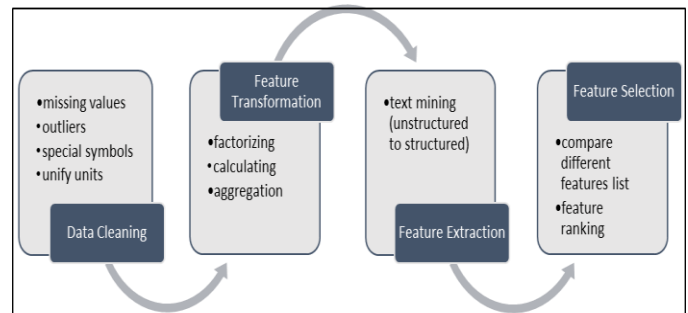| #	Risk Factor	Attributes Availability
Potentially Modifiable		
1	Low Education	Years of education
2	Hypertension	Detailed data available
3	Obesity	Can be calculated from weight\height
4	Smoking	Detailed data available
5	Depression	Detailed information
6	Diabetes	Check medical history and laboratory test results
7	Physical inactivity	Search relevant questioner's answers
8	Hearing loss	Search related terms on reported medical history
9	Social isolation	Search questioner's answers and related features (marital status, work)
10	Cognitive inactivity	Search relevant questioner's answers and related features
11	Chronic Heart disease	Check medical history
12	Alcohol use	Available
13	Chronic Kidney disease	Check medical history
Diagnosis		
14	Cognitive State	CN, MCI, and Dementia (available: baseline, follow up)

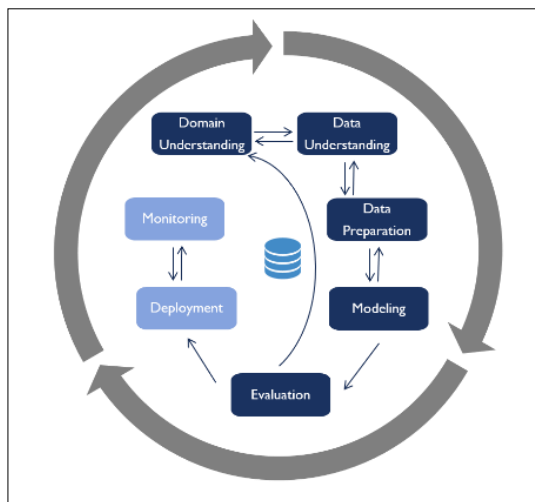Fig. 2. A Summary of the Data Preparation Steps used in this Study.

1) Dealing with missing values in longitudinal data: Based on the ADNI study description, missing data were coded with -1 or -4. Typically, -4 is used for not applicable (i.e., data is not collected at a specific visit), and -1 is used for confirmed missing data. The detailed study schedule shows the data collected at each visit for each cohort group (i.e., CN, MCI, and AD).

To check the reason for missing data and to determine whether the data were missing completely at random (MCAR), missing at random (MAR), or not missing at random (NMAR) [22] [25], the visit schedule descriptions, the visit registry table, and the exclusion tables were checked. The exclusion tables helped determine the reason for dropout, which might not be related to dementia, such as study partner availability, moving to another city, or not being willing to undergo MRI scans. For the available records missing data, several reasons were identified, and different actions were applied.

a) Missing Data Due to Scheduled Visit Design: In some cases, the data were missing because they were not collected during a visit (e.g., some visits were only for MRI imaging session; some data were collected only at the baseline). The missing data in these cases were considered MCAR and were imputed using the same patient's previous data, following the last observation carried forward (LOCF) [22] method.

- *Missing Height:* In the detailed ADNI visit schedules, participant's heights were only taken once during a screening visit, unlike their weights, which were repeatedly taken at each visit. Therefore, missing height data for each visit were filled in using a participant's screening visit height.

- *Missing Demographics and Medical History:* These data were collected only during the screening visit (repeated at the screening visit for each cohort, i.e., if a participant was included in ADNI1, 2, 3, medical history was taken at the screening visit for each cohort). The missing data were filled in using the same data for all visits (not imputation rather than fixed, although it might change, this is not recorded).

b) Missing Data Due to ADNI Study Stage Design: If the data were not collected during a specific ADNI stage, this means that the data were missing for all patients enrolled only during this stage. Therefore, only available or complete cases were considered. Examples of this include detailed smoking history, alcohol use, and medical history, which are not available for the ADNI3 study design. This led to selecting only ADNI1, ADNIGo, and ADNI2 cohorts for the study. Also, cognitive activity data were collected only beginning from ADNIGo, and thus, participants who were enrolled only in ADNI1 were excluded.

2) Feature transformation: Not all features have the desired format. Some new features must be calculated from existing ones, and some binary or categorical features must be factorized or encoded. Moreover, some features must be aggregated because they are repeated in multiple rows and could be defined as unique new features. The applied feature transformation included:

a) Unit Modifications: Height and weight units were not unified for all entries. Some were recorded as kg\cm, lbs\inch, lbs\cm, or kg\inch. All measurements were modified to the metric unit kg\cm.

b) Calculation: Some features needed to be calculated from other existing features. This may cause multicollinearity, which was reduced by selecting the best representative features which yield to better models results [11]. The calculated features were as follow:

- *BMI and Obesity:* Body mass index (BMI) was calculated based on the height and weight of the participants. Moreover, obesity was recorded when BMI >30 [27].

- *Social Isolation:* Social isolation level has been detected by calculating the available relative features, which are the marital status and retirement (as per [28], [29]).

- *Physical Activity:* Physical activity level has been calculated by adding up the related functional and physical assessment questioners' answers such as going shopping, playing games, and going out of the neighborhood.

c) Factorization: Visit codes were in a string format and were factorized to be numerical for simple computations and comparisons.

d) Aggregation: Structured medical description row-based fields were converted to binary column-based features (i.e., row for each condition per participant converted to 1 row with all conditions per participant).

e) Normalization: For modeling purpose, numerical data has been normalized to range from 0 to 1 using the MinMaxScaler.

f) Encoding Categorical Features: Categorical features were encoded using dummy variables by converting the feature of k-categories to k-1 different dummy variables [30]. This was applied to the marital status and gender variables.

3) Feature extraction: Most risk factors available within the medical history description were text entries. These descriptions were entered as a free, unstructured text field with multiple variations of the same condition, which required some preprocessing to extract the features.

Some basic text mining techniques were applied to extract the previously defined risk factors and then to check for other possible factors. Using the NLTK package, stop words were removed, the text was converted to lower case, the most common terms and n-grams were selected, word clouds were plotted, and the known risk factor terms were searched and selected. Fig. 3 illustrates how medical history descriptions differ between those with dementia and others.

After applying text mining, each unstructured medical history text field was converted to a structured field (categorized), which is illustrated by Fig. 4.
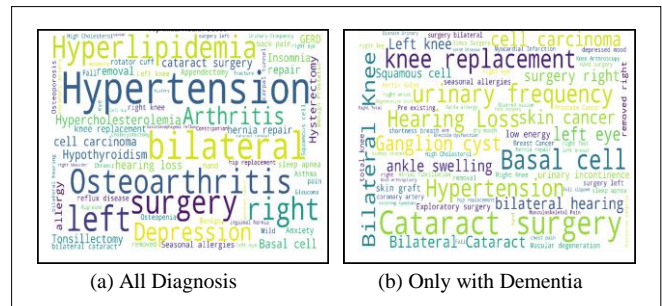

(a) All Diagnosis   (b) Only with Dementia

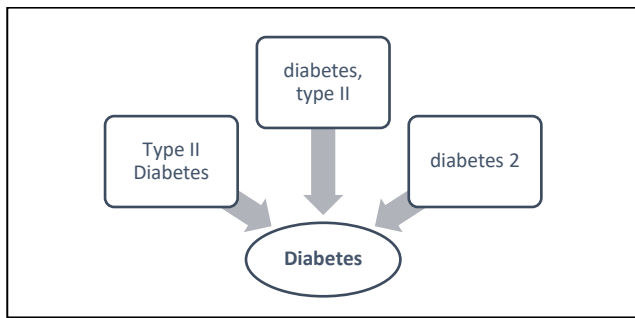Fig. 3. Word Cloud of Most Common Medical History Descriptions

Fig. 4. Unstructured to Structured Medical Descriptions.

4) Feature selection: The previously defined features that were clinically approved to be relevant were selected. Both *Lancet* commission and *Libra* index modifiable risk factors features were considered in order to check, which gives better results.

5) Data integration: All selected tables were integrated and merged into a single table with all considered features.

C. Modeling

This research focused on interpretable modeling because of its importance for informing clinicians managing patients. Interpretable machine-learning classification models, such as the Logistic Regression, Naïve Bayes, Decision Tree, and Random Forest. Both binary (dementia vs. non-dementia), and multi-class (CN vs. MCI vs. dementia) classifications were applied using the models.

1) Logistic regression: Logistic regression (LR) is an extension of linear regression and is used to solve classification problems. Basically, it is designed to solve binary classification problems where there are only two outcomes, but eventually, it is extended to support multi-classification, which is referred to as a multinomial logistic regression [26] [18]. A well-known method used to achieve a multinomial classification is using a set of one-versus-all models. For example, if there are n targets levels, n numbers of one-versus-all logistic regression models are created, and each model distinguishes between the features of one target level and all the others [26] [30].

2) Naïve bayes: In machine learning, the Naïve Bayes (NB) method serves as a probabilistic classifier that uses the Bayes' theorem of conditional probabilities [18] [26]. It assumes a strong (naïve) independence between features and calculates the class probabilities for each feature independently. The conditional probability of a class is the normalized class probability times the probability of each feature given by a class [18].

3) Decision tree: A decision tree (DT) is a tree-based model that splits the data repeatedly according to specific cutoff values in the features [18] [26]. Different subsets are created through splitting, separating instances to belong to one subset. The intermediate subsets are the internal nodes, while the final subsets are the leaf nodes. Decision trees are most useful if the relationship between the features and the target are nonlinear or if there are interactions between features.

4) Random forest (ensemble learning): The random forest (RF) model is an ensemble learning model that combines bagging, subspace sampling, and decision trees to create a more powerful model [26] [30]. The random forest model overcomes the overfitting problem of a decision tree, which is why it usually performs better.

A random forest model is a collection of decision trees in which each tree is slightly different from another. Once each individual decision tree model has been created (bagging), the ensemble makes predictions by returning the majority vote of the classifiers. This reduces the overfitting amount by averaging the results while maintaining the predictive power of each tree [30].

D. Evaluation

After the models are developed, the results were evaluated using multiple metrics and techniques to identify possible problems with overfitting and parameter tuning issues.

1) Confusion Matrix-Based performance measures: A confusion matrix is a convenient method used to comprehensively describe the performance of classification evaluations, which can be either binary or multi-class [26] [30] [31]. Most other metrics are derived from the basic components of the confusion matrix, which are the *True Positive (TP), True Negative (TN), False Positive (FP),* and *False Negative (FN),* and their percentage conversions. From these components, main evaluation measures such as accuracy, precision, recall, and F-score were calculated [31]. In this study, recall (sensitivity) is defined as the proportion of subjects who have dementia that are correctly classified. Precision is defined as the proportion of subjects who did not have dementia that are correctly classified. Accuracy is defined as the proportion of all subjects that are correctly classified, while F1 is the weighted average of precision and recall.

2) Sensitivity, specificity, and AUROC: The receiver operating characteristic (ROC) evaluates a model's true performance while considering all possible probability cutoffs (thresholds). The default threshold is 0.5; however, it could range from 0 to 1, and the classification results may change accordingly. The area under the ROC (AUROC) summarizes thresholds changes of both TPR (sensitivity) and FPR (1-specificity). The perfect fit is 1, the worst is 0, and the random prediction is 0.5.

IV. EXPERIMENTS AND RESULTS

The experiments and analysis conducted for this research were applied using the following environments, tools, and libraries:

1) Environments Used: Python (3.6.4) and R.

2) Tools Used: Jupyter Notebook version 5.4.0, Google Colab (for faster modeling), and SPSS version 24 (for missing data mechanism and quickly find and explore).

3) Main Libraries and Packages: scikit-learn (for machine learning), NLTK (for text mining).

A. Model Validation

A balanced train-and-test split was applied using the *StratifiedKFold* to split the data once into a 75% training set and a 25% testing set. This ensures the same percentage of each class per group. Moreover, longitudinal data grouping by the participant was performed using the *GroupKFold*, which is a special variant of cross-validation that takes into account the repeated measurements from the same subject and considers them as grouped data. Parameter tuning was achieved using nested cross-validation by applying *GridSearchCV* parameter tuning (inner loop) to cross-validation (outer loop).

B. Feature Selection

Using only statistically significant features either from the univariate analysis or per models was not sufficient because it decreased the accuracy from an average of 70% to 50%. This could be explained because there may be interactions between the features. The best feature selection was obtained using both the *Lancet* and the *Libra* index features. Using *Lancet* features alone gives an average of 59% accuracy only.

On the other hand, using *Libra* features only gives an average of 68% accuracy meaning that it is more comprehensive and predictive to the machine learning models, although combining it with *Lancet's* gives better results.

C. Cross-Sectional vs. Longitudinal Data Evaluation

Longitudinal data perform better than cross-sectional (baseline) data. However, the latest visit data gives the best results among them all.

Table III summarizes the results of the multi-class classifications on different data subsets for all models using ten-fold cross-validation.

TABLE. III. EVALUATION RESULTS SUMMARY (MULTI-CLASS)

Model	LR	NB	DT	RF
Longitudinal				
Accuracy	68.13%	57.99%	66.75%	**68.55%**
Precision	68.49%	65.38%	67.05%	**68.92%**
Sensitivity (Recall)	68.13%	57.99%	66.75%	**68.55%**
F1	68.18%	53.82%	66.75%	**68.61%**
Cross-Sectional (Baseline)				
Accuracy	**63.71%**	54.01%	60.34%	63.29%
Precision	64.39%	57.52%	60.30%	**65.60%**
Sensitivity (Recall)	**63.71**%	54.01%	60.34%	63.29%
F1	**63.11%**	47.39%	60.29%	61.92%
Latest Visit				
Accuracy	**77.00%**	66.77%	70.29%	71.57%
Precision	**76.76%**	66.13%	70.18%	70.73%
Sensitivity (Recall)	**77.00%**	66.77%	70.29%	71.57%
F1	**76.35%**	66.38%	70.15%	70.51%

As shown in the table, for longitudinal data, the best performance results are obtained by RF and reached around 68%. Moreover, using the baseline data alone, best results reached around 63-65% only.

Furthermore, using the latest visit data, the best performance results are obtained using LR and reached around 77%, which is the best among all data subsets. The overall differences between metrics are relatively small for all models.

Below figures illustrate the models' results of each longitudinal, baseline, and latest visit subsets respectively. The NB gives the least performance results for all data subsets.

For both longitudinal and baseline data, the LR and RF results are very similar for all metrics, followed by the DT, while the NB results are very lower, as shown in Fig. 5 and Fig. 6.

Fig. 7 shows the latest visit results, where it is clear how the LR outperformed the other models for all metrics. The DT and RF results for this subset are relatively similar to each other.

As shown, considering only the latest visit subset gives better evaluation results for all models, followed by the longitudinal, and finally, the baseline subset. This performance difference is clearly illustrated in Fig. 8 for the LR model.
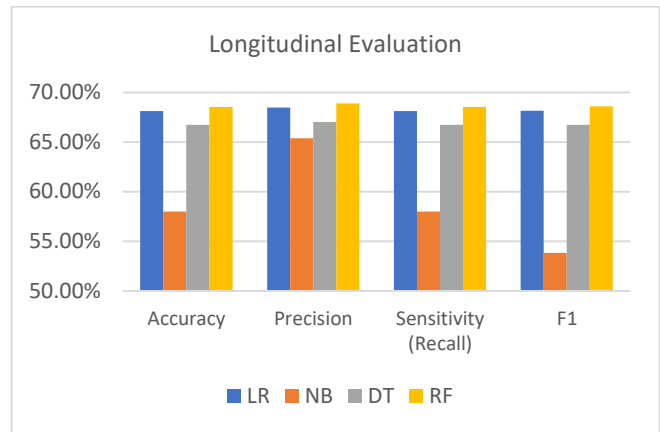

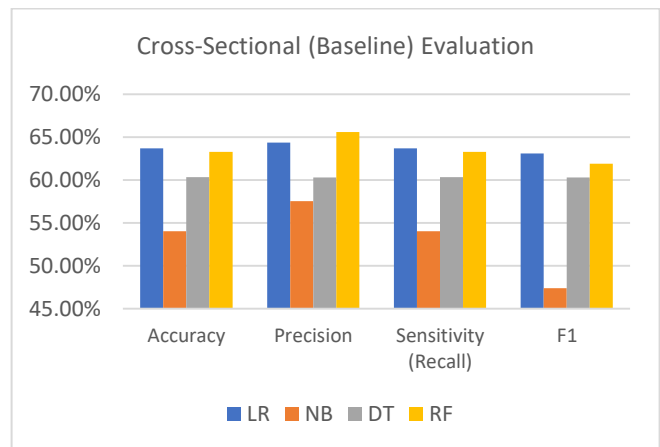Fig. 5. Longitudinal Evaluation - (Multi-Class).

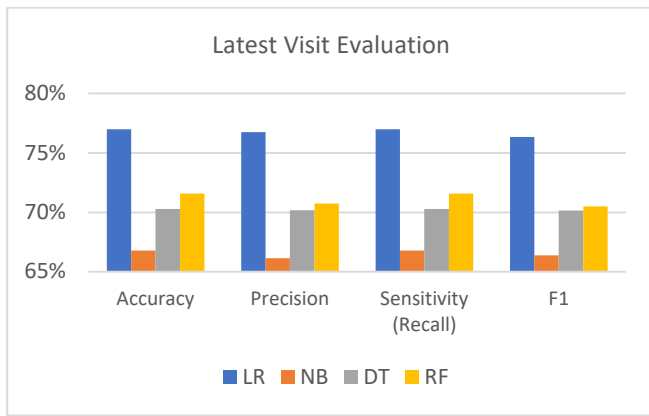Fig. 6. Cross-Sectional (Baseline) Evaluation - (Multi-Class).

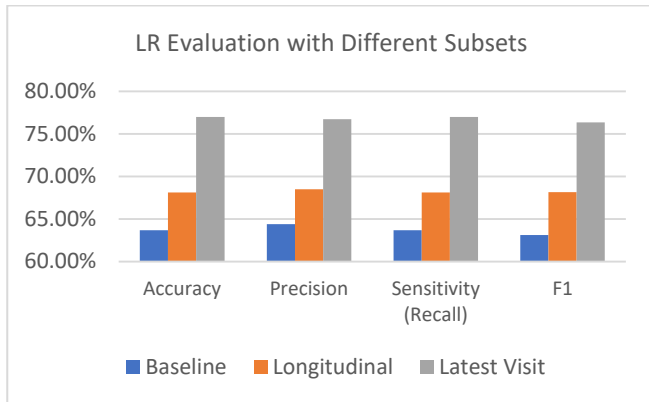Fig. 7.   Latest Visit Evaluation-(Multi-Class).

Fig. 8.   LR Evaluation Comparison of different Subsets-(Multi-Class).

The area under the ROC curve (AUROC) of the dementia class was 96% for both top models (LR and RF) as illustrated in Fig. 9 and Fig. 10, respectively. Although the AUROC of MCI and CN classes are lower than the dementia class (CN: 86%-88% and MCI: 67%-78% for LR and RF, respectively), the AUROC of the dementia is the more important in this study.

D. Binary vs. Multi-Class Evaluation

In addition to the multi-class classification (CN vs. MCI vs. dementia), binary classification (dementia vs. non-dementia) has been applied using all previous models. Binary classifications outperformed multi-class classification, although they are less informative. Table IV and Fig. 11 demonstrates the results for both the binary and the multi-class classifications for the two top performed models (i.e., LR and RF). While the multi-class models' result reaches 70% only, the binary models reach around 92% on all metrics. The LR results are better than the RF for all metrics.

E. Overfitting Check and Model Generalization

All models have been checked against overfitting by comparing the training and testing accuracies. The difference between the train and test accuracies ranged between 0 (LR and NB), 0.02 (RF), and 0.04 (DT), which is considered small and acceptable. Moreover, to ensure cross-validation generalization, the standard deviation of the accuracy for all

folds has been checked. This ranged between 0.02 and maximum 0.03, which is all considered small and acceptable. Table V shows the detailed results of the models check.
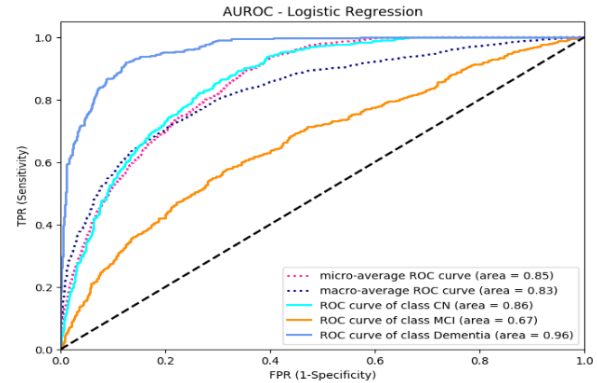
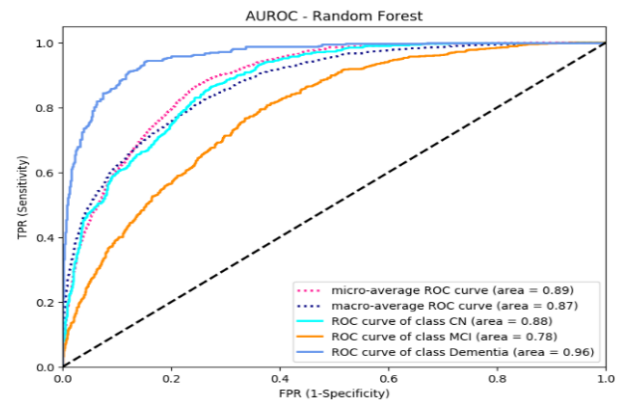Fig. 9.   AUROC for LR–Multi-Class Classification (Longitudinal).

Fig. 10. AUROC for RF–Multi-Class Classification (Longitudinal).

TABLE. IV.   EVALUATION RESULTS OF BINARY VS. MULTI-CLASS

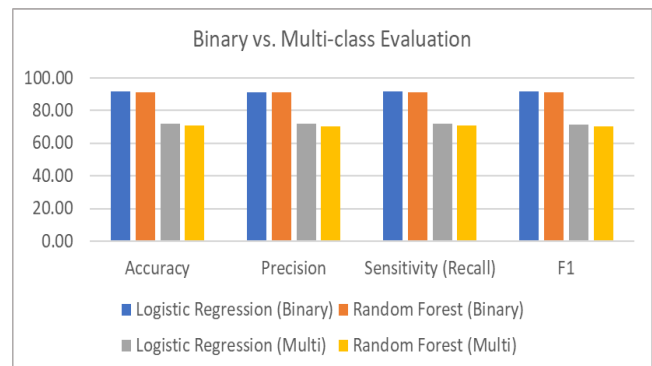| Model	LR	RF	LR	RF
	Binary		*Multi-class*	
Accuracy	**91.53%**	91.24%	77.00%	71.57%
Precision	**91.34%**	90.95%	76.76%	70.73%
Sensitivity	**91.53%**	91.24%	77.00%	71.57%
F1	**91.41%**	91.01%	76.35%	70.51%

Fig. 11. Binary vs. Multi-Class Evaluation.

TABLE. V.   OVERFITTING AND MODELS GENERALIZATION CHECK

Results\ Model	LR	NB	DT	RF
Train-Test Split				
Training Accuracy	0.92	0.87	0.94	0.93
Testing Accuracy	0.92	0.87	0.90	0.91
Training - Testing Accuracy	**0.00**	**0.00**	**0.04**	**0.02**
Cross-Validation				
CV Folds Accuracy Mean	0.91	0.87	0.88	0.90
Standard Deviation	**0.02**	**0.03**	**0.02**	**0.03**

F. Feature Importance

Feature importance was calculated using different models. For linear models, it was calculated based on the absolute values of the coefficients. For tree-based models, it was calculated based on the model's feature importance. Table VI summarizes the risk factors' importance for each model. From the table, it is clearly shown that BMI, Cognitive Activity, and Physical Activity feature importance are top across most models. The feature importance was extracted from the best performing models which combine both *Lancet* and *Libra* lists.

G. Evaluation Summary

The best classification results were obtained using both the *Lancet* and the *Libra* risk factor lists, considering the longitudinal data set which outperformed the cross-sectional baseline one. Moreover, using data of the most recent visits only provided even better results than using the whole longitudinal set.

In some cases, it is important to detect whether a person has dementia or not, while in other cases, the exact cognitive state is needed. Therefore, both binary and multi-class classifications have been applied. The binary classification yielded to about 92% accuracy, while the multi-class classification yielded to a 77% accuracy using logistic regression, followed by random forest with 92% and 70%, respectively. The area under the ROC of the dementia class was nearly perfect at 96% for both models.

TABLE. VI.   FEATURE IMPORTANCE FOR TOP MODELS

#	Feature\ Model	Feature Importance Order				
		LR	NB	DT	RF	All Models
1	BMI	3	7	3	3	16
2	Cognitive Activity	1	13	2	1	17
3	Physical Activity	2	12	1	2	17
4	Smoking	12	1	4	5	22
5	Alcohol	4	2	9	11	26
6	Heart	5	4	12	8	29
7	Kidney	6	3	7	13	29
8	Depression	8	9	8	6	31
9	Hearing Loss	9	6	6	12	33
10	Education	11	14	5	4	34
11	Hypertension	7	10	10	10	37
12	Diabetes	13	5	14	9	41
13	Social Isolation	14	11	11	7	43
14	Cholesterol	10	8	13	14	45

Although features importance was not identical for all models, the top three features importance were identical for the two top performed models (i.e., LR and RF), which is a sign of model's stability. Furthermore, as this is an observational study analysis, the feature importance of each model does not claim any causality of dementia or MCI. The importance derived from the available data may not be representative of a wider population.

The best obtained results of this study were either competitive or even better than the results obtained by other previous studies which used MRI data and machine learning or deep learning methods [11], [12], [13]. Their best overall accuracies range between 65% and 92%. Moreover, the results of this study were better than a previous study that used a machine learning approach with modifiable risk factors, where their best accuracy reaches 75.24% only [15]. However, this is not considered as a complete comparison as the other studies used different datasets.

V.   CONCLUSION

A. Achievements of the Research Objectives

The research discussed and evaluated in the previous sections aims to use different interpretable machine-learning classification models to detect dementia based on its modifiable risk factors only. It explored and applied *Lancet,* and *Libra* lists of modifiable risk factors on the ADNI dataset, which is as far as known have not been applied on this dataset using machine learning approaches.

The best classification results were obtained using both the *Lancet* and the *Libra* risk factor lists. Considering the longitudinal data set outperformed the cross-sectional baseline one. Moreover, using data of the most recent visits only provided even better results than using the whole longitudinal set.

The binary classification yielded to about 92% accuracy, while the multi-class classification yielded to a 77% accuracy using logistic regression, followed by random forest with 92% and 70%, respectively. Furthermore, the best achieved overall accuracies were either competitive to or better than previous studies results.

B. Limitations

This research involved an experimental analysis of an observational study based on the ADNI dataset, and there is no claim to present causations. The ADNI study was not primarily designed to address the modifiable risk factors; thus, it may lack some useful features, especially during the early and middle life courses. Social isolation and physical activities are not explicitly addressed by the study, and the results may be more accurate if more detailed data for these factors were collected. Medical history and other important useful demographic features, such as occupation, were collected as free text and were not categorized in a structured format during the data collection stage, which may have helped make the analysis simpler and more accurate.

REFERENCES

[1] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, et al., "Dementia prevention, intervention, and care," The Lancet, vol. 390, pp. 2673-2734, 2017.

[2] A. P. Association, Diagnostic and statistical manual of mental disorders (DSM-5®): American Psychiatric Pub, 2013.

[3] S. J. Vos, M. P. Van Boxtel, O. J. Schiepers, K. Deckers, M. De Vugt, I. Carriere, et al., "Modifiable risk factors for prevention of dementia in midlife, late life and the oldest-old: validation of the LIBRA Index," Journal of Alzheimer's Disease, vol. 58, pp. 537-547, 2017.

[4] K. Yaffe, "Modifiable risk factors and prevention of dementia: What is the latest evidence?," JAMA Internal Medicine, vol. 178, pp. 281-282, 2018.

[5] K. Deckers, M. P. van Boxtel, O. J. Schiepers, M. de Vugt, J. L. Muñoz Sánchez, K. J. Anstey, et al., "Target risk factors for dementia prevention: a systematic review and Delphi consensus study on the evidence from observational studies," International journal of geriatric psychiatry, vol. 30, pp. 234-246, 2015.

[6] O. J. Schiepers, S. Köhler, K. Deckers, K. Irving, C. A. O'donnell, M. van den Akker, et al., "Lifestyle for Brain Health (LIBRA): a new model for dementia prevention," International journal of geriatric psychiatry, vol. 33, pp. 167-175, 2018.

[7] C. A. O'Donnell, V. Manera, S. Köhler, and K. Irving, "Promoting modifiable risk factors for dementia: is there a role for general practice?," The British journal of general practice : the journal of the Royal College of General Practitioners, vol. 65, pp. 567-568, 2015.

[8] E. National Academies of Sciences and Medicine, Preventing cognitive decline and dementia: A way forward: National Academies Press, 2017.

[9] S. Tariq and P. A. Barber, "Dementia risk and prevention by targeting modifiable vascular risk factors," Journal of Neurochemistry, vol. 144, pp. 565-581, 2018/03/01 2018.

[10] X.-H. Hou, L. Feng, C. Zhang, X.-P. Cao, L. Tan, and J.-T. Yu, "Models for predicting risk of dementia: a systematic review," Journal of Neurology, Neurosurgery, and Psychiatry, 2018.

[11] B. Bratić, V. Kurbalija, M. Ivanović, I. Oder, and Z. Bosnić, "Machine Learning for Predicting Cognitive Diseases: Methods, Data Sources and Risk Factors," Journal of Medical Systems, vol. 42, p. 243, 2018/10/27 2018.

[12] R. Casanova, F.-C. Hsu, K. M. Sink, S. R. Rapp, J. D. Williamson, S. M. Resnick, et al., "Alzheimer's Disease Risk Assessment Using Large-Scale Machine Learning Methods," PLOS ONE, vol. 8, p. e77949, 2013.

[13] R. Casanova, R. T. Barnard, S. A. Gaussoin, S. Saldana, K. M. Hayden, J. E. Manson, et al., "Using high-dimensional machine learning methods to estimate an anatomical risk factor for Alzheimer's disease across imaging databases," NeuroImage, vol. 183, pp. 401-411, 2018/12/01/ 2018.

[14] M. Baumgart, H. M. Snyder, M. C. Carrillo, S. Fazio, H. Kim, and H. Johns, "Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective," Alzheimer's & Dementia, vol. 11, pp. 718-726, 2015/06/01/ 2015.

[15] J. O'Donoghue, M. Roantree, and A. McCarren, "Variable interactions in risk factors for dementia," in 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS), 2016, pp. 1-10.

[16] S. Joshi, P. D. Shenoy, K. Venugopal, and L. Patnaik, "Classification of neurodegenerative disorders based on major risk factors employing machine learning techniques," International Journal of Engineering and Technology, vol. 2, p. 350, 2010.

[17] Y. Ding, J. H. Sohn, M. G. Kawczynski, H. Trivedi, R. Harnish, N. W. Jenkins, et al., "A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain," Radiology, p. 180958, 2018.

[18] C. Molnar, Interpretable machine learning: A guide for making black box models explainable: Leanpub, 2019.

[19] P. Diggle, P. Heagerty, K.-Y. Liang, and S. Zeger, Analysis of Longitudinal Data vol. 25: OUP Oxford, 2013.

[20] S. Menard, Handbook of longitudinal research: Design, measurement, and analysis: Elsevier, 2007.

[21] C. R. Jack Jr, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, et al., "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 27, pp. 685-691, 2008.

[22] M. Nakai and W. Ke, "Review of the methods for handling missing data in longitudinal data analysis," International Journal of Mathematical Analysis, vol. 5, pp. 1-13, 2011.

[23] F. Ben Bouallègue, D. Mariano-Goulart, P. Payoux, and I. Alzheimer's Disease Neuroimaging, "Comparison of CSF markers and semi-quantitative amyloid PET in Alzheimer's disease diagnosis and in

cognitive impairment prognosis using the ADNI-2 database," Alzheimer's research & therapy, vol. 9, pp. 32-32, 2017.

[24] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, and A. s. D. N. Initiative, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," Neuroimage, vol. 55, pp. 856-867, 2011.

[25] J. M. Engels and P. Diehr, "Imputation of missing longitudinal data: a comparison of methods," Journal of Clinical Epidemiology, vol. 56, pp. 968-976, 2003/10/01/ 2003.

[26] J. D. Kelleher, B. Mac Namee, and A. D'arcy, Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies: MIT Press, 2015.

[27] G. Piumatti, S. C. Moore, D. M. Berridge, C. Sarkar, and J. Gallacher, "The relationship between alcohol use and long-term cognitive decline in middle and late life: a longitudinal analysis using UK Biobank," Journal of Public Health, vol. 40, pp. 304-311, 2018.

[28] E. Y. Cornwell and L. J. Waite, "Measuring social isolation among older adults using multiple indicators from the NSHAP study," The journals of gerontology. Series B, Psychological sciences and social sciences, vol. 64 Suppl 1, pp. i38-i46, 2009.

[29] J. Holt-Lunstad, T. B. Smith, M. Baker, T. Harris, and D. Stephenson, "Loneliness and Social Isolation as Risk Factors for Mortality: A Meta-Analytic Review," Perspectives on Psychological Science, vol. 10, pp. 227-237, 2015/03/01 2015.

[30] A. C. Muller and S. Guido, Introduction to machine learning with Python: a guide for data scientists: O'Reilly Media, 2017.

[31] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, pp. 427-437, 2009.