

Review

Not peer-reviewed version

Predictive Analytics in Heart Failure Prediction: A Systematic Review

Qisthi Alhazmi Hidayaturrohman and [Eisuke Hanada](#) *

Posted Date: 24 January 2024

doi: 10.20944/preprints202401.1671.v1

Keywords: heart failure; predictive analytics; machine learning; readmission; mortality



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Predictive Analytics in Heart Failure Prediction: A Systematic Review

Qisthi Alhazmi Hidayaturrohman ^{1,3} and Eisuke Hanada ^{2,*}

¹ Graduate School of Science and Engineering, Saga University; qisthialhazmi@upnvj.ac.id; <https://orcid.org/0000-0002-8615-9076>

² Faculty of Science and Engineering, Saga University; hanada@cc.saga-u.ac.jp; <https://orcid.org/0000-0002-6718-0101>

³ Faculty of Engineering, Universitas Pembangunan Nasional Veteran Jakarta;

* Correspondence: hanada@cc.saga-u.ac.jp

Abstract: Heart failure is a leading cause of death among people worldwide. The cost of treatment can be prohibitive, and early prediction of heart failure would reduce treatment costs to patients and hospitals. Improved prediction of readmission would also be of great help hospital, allowing them to better manage their treatment programs and budgets. This systematic review aims to summarize recent studies of predictive analytics models that have been constructed to predict heart failure and readmission. Some of these models use a statistical approach and others a machine learning approach. Electronic patient health records, including demography, physical and clinical values, and laboratory findings are used to design and build predictive models. Predictive analytics with good performance have become essential for clinicians and specialists. Choosing a suitable machine learning algorithm and data preprocessing technique will improve predictive model performance, including data imputation and class imbalance.

Keywords: heart failure; predictive analytics; machine learning; readmission; mortality

1. Introduction

Heart disease is one of the deadliest diseases among people worldwide [1], with 50% of heart failure patients dying within five years [2]. Doctors use a variety of tests to diagnose heart failure, including physical examination, blood and laboratory tests, and family history of the disease [1]. Even though there is no cure for heart failure, delicate medical procedures and treatments improve quality of life [2]. Researchers have conducted various studies to build prediction models to assist with the diagnosis of heart failure: early diagnosis allows a patient to get the proper treatment and minimizes the seriousness of this disease.

Medical procedures and treatments are expensive, especially for hospitalized patients [3]. To help reduce the expense, researchers have done numerous studies to identify heart failure patients who will need readmission. As information technology has developed, especially artificial intelligence, researchers have begun developing better early prediction systems. Including machine learning in the models improves their performance and prediction quality. With the right dataset and suitable data processing techniques, a predictive model's performance should improve [4]. Predictive analytics has become the most used approach by researchers who construct predictive models using machine learning. In this approach, researchers use data gleaned from electronic health records to predict hospital readmission and mortality. However, there are serious obstacles to the development of predictive models that use a predictive analytics approach. Problems with data presentation and addressing problems such as imbalance in data class create challenges for researchers.

In this literature review, we provide an overview of and explanation of how predictive analytics can be used to predict heart failure, readmission, and mortality among patients with heart failure.

2. Materials and Methods

2.1. Literature Exploration

Before starting our search for research articles on predictive analytics in heart failure, it was necessary to define the targets, topics, and themes necessary for predictive analytics and heart failure prediction: five such targets were defined. Furthermore, we chose closely related keywords that would be critical to our Google Scholar and PubMed search. We limited our search to the years from 2010 to 2023 to include only the most recent studies conducted in this field. Quotation mark- and Boolean operators were used to search for titles and abstracts that were closely related to the defined topic. After doing this wide search, the next step was to select the studies to be included.

2.2. Study Selection and Screening

In the study selection step, several considerations went into the pairing down the articles found in the preliminary search, as follows:

1. A paper must have been published in a journal or conference booklet to be selected. Books, book series, chapters, and others were not considered.
2. It must be a research paper, not a review, a meta-analysis, or a literature review.
3. We considered the credibility and quality of the publisher. To do this, we cross-checked the publisher and journal with Scimago/Scopus and Clarivate or the Web of Science.
4. The final consideration is that the published papers should include the full text. Our university has limited access to journal subscriptions, which limited us somewhat in selecting the papers.

To eliminate papers that varied from the topic and to select those that met our conditions, we read the full titles and abstracts of the selected papers, paying careful attention to the limits we set, which included that they contain machine learning algorithms, use a designated dataset, and have data processing involvement. This was important to ensure that the paper was related to heart failure prediction or the prediction of readmission. Papers meeting the above conditions in the screening stages were extracted for this review.

2.3. Data Extraction

The items extracted included the author's name, year of publication, study objective, origin of data, dataset specifications, machine learning algorithm(s), methodology, and the evaluation of the model.

2.4. Literature Review Diagram Flow

In the search stages, combining keywords with quotation mark- and Boolean operators identified 3,270 publications. After applying the full set of conditions specified above, 65 papers were selected for this literature review. The diagram in Figure 1 below shows the flow of selection.

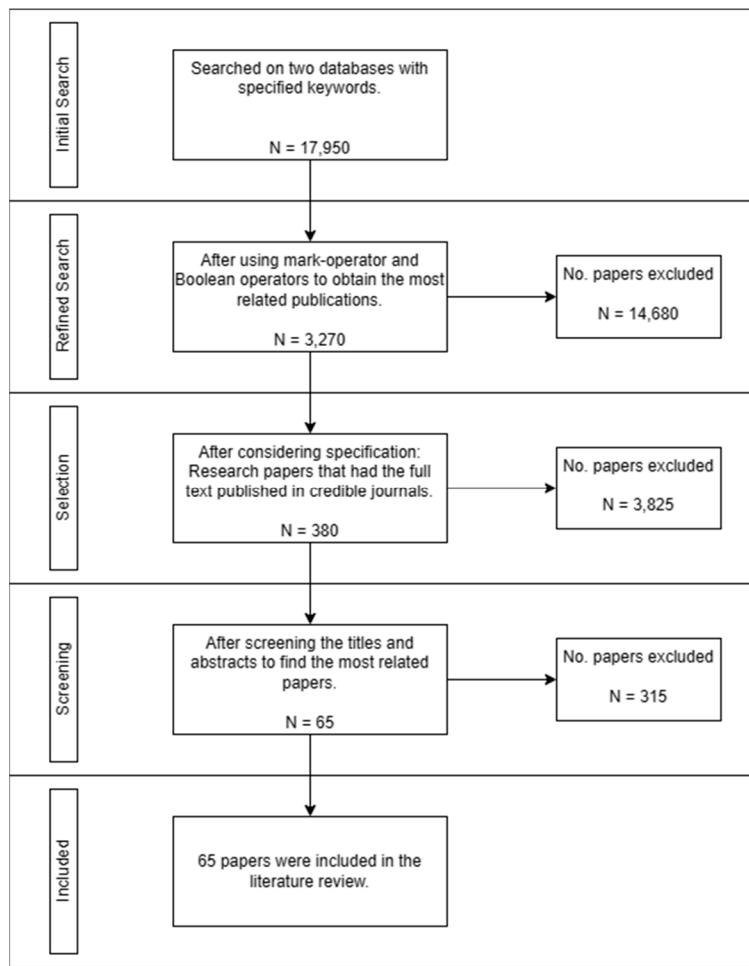


Figure 1. Literature review flow diagram.

2.5. Classification of the Paper

We analyzed the 65 papers to be reviewed comprehensively by classifying them into two categories. The first category includes 26 papers that mainly aim to predict or diagnose heart failure or risk of heart failure using statistical or machine learning approaches to build a predictive analytics model and that do heart failure prediction by building predictive analytics models, using either their own or open-accessed datasets. The second category includes 39 papers that aimed to predict readmission or mortality among patients with heart failure. They were divided into three subcategories: readmission, mortality, or both. Seventeen publications mainly discussed implementing predictive analytics models to predict rehospitalization and 14 discussed mortality prediction. The other eight discussed both readmission and mortality prediction.

3. Results

3.1. Predictive Analytics of Heart Failure Prediction

We found in these recent studies that artificial intelligence, specifically machine learning, plays an important role in predicting heart failure and risk from electronic medical records. This approach will be of great help to the clinical decision-making process and in diagnosing patients with heart failure. Study [5], designed a clinical decision support system (CDSS) by implementing machine learning in decision-making to evaluate the severity of HF among patients with HF. Machine learning algorithms were involved in building the predictive models of the study and in evaluating all of the models by performing cross-validation of each model. The study concluded that machine learning-

based CDSS is useful for diagnosing heart failure and is readable even when done by non-cardiologists or non-clinician users.

Involving more than 400,000 primary care patients, studies [6] and [7] used a machine learning algorithm to diagnose heart failure in primary care patients collected by the Geisinger Clinic. These studies were able to predict heart failure in different time windows and showed good performance. The models in [6] used unstructured and structured data in model building, while the models in [7] were built by processing EHR data. Similar to those studies, [8] used the data of real patients at King Saud University Medical City (KSUMC) and manually extracted the necessary information of the patients to be processed to predictive models using a machine learning approach in a big data environment. The study involved PCA's pre-processing and feature reduction techniques to obtain a promising predictive result. The study performed well in predicting heart failure, although the number of patients was only 100.

Aside from machine learning in building its predictive models, study [9] involved more than one million elderly patients collected by Medicare USA and built predictive models using a trajectory-based disease progression model to predict heart failure among unseen patients. [10] also implemented the Cox hazard proportional model to predict the risk of heart failure by patients collected by COOL-AF Thailand between 2014 and 2017. The study evaluated their predictive model by calculating the model's C-index, D-statistics, calibration plot, brier test, and survival analysis. The proposed model provided good prediction of heart failure.

Study [11] described using a predictive analytics approach to predicting Heart Failure with preserved Ejection Fraction (HFpEF), a subtype of heart failure. This study built predictive models using five different machine-learning algorithms. It assessed those models using c-statistic, brier score, sensitivity, and specificity in the performance evaluation of the predictive models. The study stated that predictive analytics accurately predicted the presence of HFpEF in patients with heart failure. Similar to the previous model, [12] used three algorithms to predict and identify Acute Decompensated Heart Failure (ADHF) in patients collected by Tisch Hospital, USA. It assessed each model with AUC, sensitivity, and PPV to determine which model performed best. The study found that a machine learning-based predictive model best predicted ADHF.

Study [13] used datasets provided by the University College of Dublin (UCD) and the Department of Cardiology of the Hospital University Ioannina that included 487 patients segregated by type of heart failure. The study implemented a feature for removal of data with more than or equal to 50% missing values, removed discrete features with unbalanced distribution, and detected and corrected the outliers and typos of the dataset. The study also performed class balancing in the next step by applying an under-sampling technique to the dataset. Through this approach they were able to obtain an ideal dataset. It generated promising results for classifying heart failure by dividing the main dataset into a sub-dataset for each type.

Studies have shown that the predictive analytics approach can optimally predict heart failure. Furthermore, this approach can also explain how the predictive models do the prediction. [14] describes the use of model interpretation and feature importance explanation using the SHAP approach to give physicians an understanding of the models. The models of this study used five machine learning algorithms to process 5,004 data sets from the Medical University Hospital in Shanxi Province of China. It used the Shapley additive explanations (SHAP) approach for the best-performed model to interpret the model and its feature importance. To prove the effectiveness of the predictive analytics approach using machine learning, the experimental study of [15] showed that in predicting cardiovascular disease, of which heart failure is one type, predictive analytics using machine learning methods outperformed other risk scales like SCORE and REGICOR. The study was part of an analytical observational study of the ESCARVAL RISK Cohort in Spain. In this study, the machine learning-based predictive models were compared to SCORE and REGICOR to predict the cardiovascular risk of patients in the study cohort.

Various studies are cohort studies or use hospital data, as did the previously discussed studies. However, there are selected studies that build predictive models using open, public data from machine learning repositories such as Physionet and UCI. Researchers commonly use heart disease

and heart failure datasets from the UCI machine learning repository to build predictive models to predict heart failure and the risk of heart failure. [16], [17], and [18] used the Heart Disease dataset of the UCI database to build their models and involved various machine learning algorithms in their construction. They assessed each model with classification metrics like accuracy, precision, F1 score, recall, and AUC-ROC score to discover the performance of each model.

Because the Heart Disease dataset of the UCI repository has missing values, various techniques must be used to address problems associated with the missing values. Study [19] removed all the missing values from the dataset. However, this technique might result in biased prediction results. [20] imputed the missing values using the k-nearest Neighbor imputation technique. The study resulted in better performance compared to the previous study, which only removed the missing values without performing imputation.

A study by [21] showed that solving an imbalance problem might improve performance. The study used the SMOTE technique to address the imbalance in the Heart Disease dataset and built predictive models with six different algorithms. The study showed that using the balancing method before building their predictive models improved the predictive performance. However, because the dataset used had a slightly different number of each class, the imbalance correction technique was not necessary. In the case of [22], the study also used the SMOTE technique to solve the imbalance problem in the Heart Failure dataset. In the predictive model of [22], SMOTE-ENN, the SMOTE technique combined with ENN, was used to address the imbalance in the Heart Failure dataset. They did scaling and standardization techniques to normalize the dataset before addressing the imbalance. The study showed improved prediction performance compared to models without balancing and normalization. Like the previous study, [23] used the SMOTE technique and performed feature selection to obtain an ideal dataset with the most important features. The study resulted in good classification performance compared to the other studies mentioned.

Similar to [23] that performs feature selection, [24] and [25] implemented feature selection and optimization to obtain the most important features for the ideal heart disease dataset from the UCI database. [24] adopted the KS-Test to select the optimal attributes for the dataset and built a predictive model using a decision tree algorithm. However, the proposed model was not significantly improved when using Mathew's correlation test to evaluate its predictive model. In [25], the Lasso algorithm was used to select the features of the Heart Disease dataset. Compared to the previous study, this study resulted in better classification performance in predicting heart failure. Previous experimental studies pre-processed data to obtain an ideal UCI Heart Disease and Heart Failure dataset, so the predictive models resulted in good performance and prediction. [26] did feature selection using Pearson's correlation to obtain an ideal version of the UCI heart disease dataset after discarding variables with missing values. The study then named the pre-processed and ideal dataset 'Satvi'. The study shows improved predictive model performance for the UCI Heart Disease dataset.

Unlike other studies that used fundamental or ensemble machine learning in building their predictive models, [27] used a big data approach and the UCI Heart Disease dataset to analyze and predict heart status. The study implemented a clustering technique to filter unnecessary data and improve the prediction effectiveness. In this study the proposed model with a clusterization approach resulted in outstanding predictive performance with high CPU utilization and low processing time. In another study, [28], implementing quantum computing in machine learning and deep learning algorithms resulted in better predictive performance than conventional machine learning algorithms. The study solved the complex dimension and size problem of the UCI Heart Disease dataset by using quantum computing to build predictive models, then compared their classification performance with other machine learning algorithms.

Similar to the UCI repository, Physionet provides various datasets in the field of health and medicine. Study [29] used datasets from the MIT-BIH and BIDMC databases, which are publicly open to access and use through Physionet. The study aimed to predict heart failure by analyzing the ECG signals from the datasets they used. By implementing a deep learning algorithm, the study was able to build a predictive model with good predictive model performance in terms of accuracy, sensitivity, and specificity in predicting heart disease. Using the MIMIC-III datasets of Physionet, [30] was able

to predict the length of stay for patients with heart failure by implementing machine learning to build their predictive models. By using several machine learning models, the predictive analytics approach in this study was able to provide good prediction and model performance in their evaluation stage. Various studies give evidence of the benefit of implementing predictive analytics to predict heart failure and the risk of heart failure. The predictive analytics approach will be of great help to participants in the health sector, especially clinicians and physicians, by making it easier to identify and diagnose heart failure and by estimating the risk of heart failure risk at an early date.

3.2. Predictive Analytics for the Prediction of Readmission or Mortality

Patients with heart failure face high risk of hospital readmission after their initial hospitalization, and mortality from heart failure is high. As with the implementation of predictive analytics to predict heart failure and its risk, this approach can be applied to predicting the need for hospitalization and readmission and for predicting and reducing mortality in patients with heart failure.

3.2.1. Readmission

EHR data was the foundation for the predictive analytics for hospital readmission among patients with heart failure in study [31], which used the Naïve Bayes model to build their model. The study used a dataset collected by Mount Sinai Hospital in 2014. The dataset had diagnosis codes, medications, lab measurements, surgical procedures, and vital signs. The proposed model reached 0.78 in the AUC score with an accuracy of 83.19%. In contrast, most papers had an AUC score for readmission in the range of 0.6 to 0.7 [31]. Involving more than one hospital, the study of [32] developed predictive analytics models from the 52 hospitals in the Korea Acute Myocardial Infarction-National Institutes of Health (KAMIR-NIH) registry. They used data from 6, 12, and 24-month follow-ups after hospital discharge and were able to predict rehospitalization of patients with acute myocardial infarction (AMI). This study used a deep learning algorithm to build the predictive model. It resulted in better performance than traditional machine learning models.

From the medical records of other countries, [33] implemented machine learning algorithms in the building of their predictive models to predict the risk of hospitalization for patients with heart failure in Rwanda. This study involved the medical records of seven hospitals and was done between 2018 and 2019. In their model prediction assessment and evaluation, their findings showed that their approach was able to predict high risk of hospitalization, with good performance. From South East Asia, a Philippine study, [34], was able to predict readmission in a sample of 322 patients with heart failure who were collected from NMMC in Cagayan de Oro City, Philippines, between June 2017 and December 2018. They also presented risk factors associated with heart failure readmission based on the most successful predictive model proposed in their study. Study [35] built predictive models that used machine learning algorithms to predict 1- and 3-month hospitalization of patients with heart failure. This experimental study involved nurses and cardiologists from medical centers in Iran who were able to collect 230 patient records. From the data collected, the study used a data mining approach with the IBM SPSS Modeler to build predictive models that used machine learning algorithms for classification. Their study models reached up to a 0.73 AUC score for the prediction of readmission.

By utilizing a database collected by University of Utah Health Care, [36] built a predictive model based on a self-developed algorithm to predict the readmission of patients with chronic heart failure and evaluated the proposed model with c-statistic, accuracy, sensitivity, specificity, PPV, and NPV to determine its performance. By dividing the dataset used into a three-step approach, including validation, this study produced a high-performance predictive analytics model that can be applied in various healthcare areas to improve the effectiveness of patient care. In another study involving university health care, [37] used more than 1 million sets of data from the University of Virginia Clinical Database Repository (CDR) collected over 15 years. The study described using the Random Forest algorithm for processing and building a predictive model from administrative data to predict unplanned, all-cause, 30-day readmission of patients with Chronic Heart Failure (CHF). The proposed model reached 0.8 AUC and an adequate c-statistic score in their evaluation of their model's

ability to predict readmission. Similar to the previous study, [38] describes the implementation of their predictive analytics by demonstrating how their decision tree algorithm produced a transparent analysis of their prediction of readmission of patients with heart failure, acute myocardial infarction, and pneumonia. The study used a decision tree algorithm, with quite good results based on the AUC score of the model.

Study [39] proposed a new data processing approach to extracting data from medical records to improve the performance of readmission prediction among patients with chronic diseases like heart failure and COPD. Their study obtained and processed data from the Center for Health Systems Innovation (CHSI) at Oklahoma State University (OSU). By assessing the AUC-ROC value from the predictive model built from a processed and balanced combined dataset, their model was able to predict readmission of patients with chronic disease and provides a competitive advantage for predictive analytics built from medical records. The study conducted by [40] used data from Partners Healthcare System, which is connected to several medical and health centers, to obtain patient data. Their studies implemented predictive analytics for the prediction of 30-day hospital readmission of patients with heart failure. By processing structured and unstructured data, the study was able to predict 30-day readmission and performed best when using their modified deep learning algorithm.

In addition to collecting data from the hospital or health center directly, two experimental studies were done as part of long-term projects, the Tele-HF project from the USA [41] and the Gestione Integrata dello Scompenso Cardiaco (GISC) study from Italy [42]. Study [41] involved 1,653 patients collected for the Tele-HF trial within 30 days of discharge to build predictive analytics for predicting readmission of patients with heart failure. The study used different machine learnings to build their models. It assessed them with the c-statistic value, which measures model discrimination. The study showed that their proposed approach provided a 17.8% improvement in predicting 30-day all-cause readmission. The study of [42] compared various machine learning algorithms in terms of their ability to predict the hospitalization of the patients with heart failure in the GISC study. By performing three different approaches to data imputation, the proposed models demonstrated various performances in predicting readmission, based on each model's PPV, NPV, sensitivity, specificity, accuracy, and AUC-ROC value. The study reports that the proposed approaches were able to predict readmission of patients with heart failure.

Differing from other studies, those of [43] and [44] used open, public data from the MIMIC-III dataset in building their predictive analytics to predict readmission of heart failure patients. [43] used structured and unstructured data from the MIMIC-III dataset to produce predictive models based on a logistic regression model. By dividing the processed dataset into five types based on the proposed iteration, they were able to achieve high performance in accuracy and AUC score for predicting readmission. Attempting a different approach, [44] used only unstructured data, clinical notes provided by the MIMIC-III dataset, to build their predictive models based on the NLP CNN model, a deep learning approach to predicting readmission. The proposed approach reached an accuracy of more than 70% for predicting readmission.

Various scores and scaling techniques can be implemented in predictive models to verify and validate the predictive analytics approach used for predicting readmission. Study [45] developed predictive analytics models with 12 different machine learning algorithms and compared each model's performance to the LaCE score, a common method used for predicting patient readmission. By assessing the AUROC value as the discrimination performance of each model, they showed that most of their predictive analytics models with machine learning algorithms provided better performance in predicting readmission than did the LaCE score. Study [46] compared their proposed approach, the Potentially Avoidable Readmission (PAR) approach, to determine the readmission score of the model and compared it with other methods, such as 3M of the Centers for Medicare and Medicaid Services (CMS) and Potentially Preventable Readmission (PPR). The study showed that their approach was able to avoid patient readmission within two weeks after discharge from the hospital.

Involving expert cardiologists in the selection of features of the heart failure data collected from Sheba Medical Center, study [47] generated better performance in terms of a combination of accuracy,

F1 score, recall, precision, and AUC (CI 95%) in predicting 30-day readmission of patients with heart failure. The study implemented three approaches to building their predictive models, categorized by the method of feature selection: machine-based, human expert-based, and collaboration-based between machine and human features. This study provided a different perspective on developing predictive hospital readmission analytics.

3.2.2. Mortality

The ability to predict the risk of death for individual patients with heart problems would be extremely valuable to health care professionals and help reverse the trend toward high cost. By leveraging large amounts of clinical data from the Geisinger EHR, which contains the data of almost 27,000 patients with heart failure and 276,819 episodes, [48] generated a new approach to building mortality prediction models that use machine learning algorithms. The study divided the dataset into training, validation, and prediction sets. The training set was used to build their prediction models by performing a split-by-year training scheme from 2013 to 2018. A validation set was used to validate the mortality prediction performance of the proposed models, which contained an average of 548 physician contacts/patient as of January 1, 2018. Their best performing predictive model was used to predict mortality for living patients with heart failure. This study presented good accuracy in predicting all-cause mortality in this large cohort. In a study done between 2013 and 2019 that involved data from 70 Japanese hospitals that contribute to the Tokushukai Medical Database, which contains the data of 1,416 patients, [49] generated machine learning-based predictive models to predict in-hospital mortality, initiation of acute renal replacement therapy, and mechanical ventilation for intensive care unit admitted patients with acute heart failure. By proposing three types of data; static, time-series, and a combination; the study produced varying prediction performance in terms of F1, recall, precision, ROC AUC, and PR AUC, especially in the prediction of in-hospital mortality.

Study [50] generated a new mortality risk predictive analytics model named MARKER-HF. It uses the electronic medical records of 5,822 patients collected by the University of California San Diego (UCSD). Their model used 3,836 UCSD data sets and was validated with the data of 1,989 patients in the UCSD dataset, 1,516 in the University of California, San Francisco (UCSF) dataset, and 888 in the BOSTAT-CHF dataset. The proposed model was able to predict the mortality risk separately for high- and low-risk groups based on the MARKER-HF score and by evaluating the AUC and sensitivity of the model. Like the previous study, [51] compared seven different mortality prediction models to predict the inpatient mortality of hospitalized patients with acute decompensated heart failure collected from the HealthFacts data of Cerner Corporation. This study evaluated seven previously proposed approaches to building a mortality predictive model from various datasets. It is notable that this study did not build predictive models with its own approach but used previously published approaches to build the models using datasets different from the published studies. This study showed that the seven proposed approaches tested performed well in predicting mortality in terms of discrimination, calibration, and specificity. Study [52] showed that the Seattle Heart Failure Model (SHFM) could predict mortality using the Mayo Clinic dataset, which contains the data of 119,749 patients collected from 1993 to 2013. The study generated their own proposed approach to building a prediction model, and their results showed better performance than the SHFM approach, with improved prediction accuracy compared to the SHFM approach.

Studies that use open, public data, like the Heart Failure dataset from the UCI repository and MIMIC-III or MIMIC-IV from Physionet, benefit researchers in developing and building mortality risk predictive analytics models. In [53], predictive analytics models were built based on machine learning algorithms to predict the survival possibility in a sample of patients with heart failure. The study used the UCI heart failure dataset in building their models and evaluated them by their accuracy, precision, recall, and F1 score. The Heart Failure dataset labels patients as either dead or alive, which allowed the proposed models to predict mortality. Like the above study, [54] used various machine learning algorithms to build mortality prediction models based on the UCI heart failure dataset. Unlike [53]'s approaches, this study used the PCA method to determine the

effectiveness of dimension reduction in prediction performance. The results of this study showed that performance was improved in terms of classification accuracy.

Using the same dataset, study [55] used a machine learning approach to determine the survival possibility of heart failure patients based on the UCI heart failure dataset. Because the dataset contains an imbalanced class, it used the SMOTE technique to address the imbalance problem. By evaluating their accuracy, precision, recall, and F1 score, the proposed models resulted in better prediction performance than studies [53] and [54]. However, [56] claimed that the oversampling technique might cause a loss of information when addressing the imbalance problem. This study developed a robust Random Forest classifier that could handle imbalance problems in the UCI heart failure dataset. With their BRF, this study was able to predict mortality using a UCI heart failure dataset, with fine prediction performance.

Studies [57] and [58] similarly used special selection methods to find the most important features of the UCI heart failure dataset. However, in [57] the feature selection method of recursive feature elimination reduced the prediction performance, so they built mortality predictive models without the feature selection method. In contrast, study [58] showed improvement in predicting mortality. It used both machine learning-based feature selection and biostatistical-based feature selection to identify the most important features in the dataset. Predictive analytics models with the most important features showed better performance in terms of Matthews Correlation Coefficient (MCC), F1 score, accuracy, and AUC ROC.

Leveraging the MIMIC-III dataset in Physionet, which is available to the public, study [59] developed predictive analytics models for all-cause in-hospital mortality of ICU-admitted patients with heart failure. Consisting of 30,000 ICU admitted patients, the MIMIC-III provided 13,389 cases of heart failure with a possible outcome of readmission or mortality. With the data of 1,177 patients after the pre-processing stage, the proposed models could predict mortality with good performance in terms of AUC-ROC and calibration c-statistic test through the machine learning algorithms used to build the predictive models. Like the previous experiment, [60] used the MIMIC-III dataset to generate predictive analytics models for the prediction of in-hospital mortality among ICU-admitted heart failure patients. Both studies implemented the same machine algorithm, XGBoost, to build their predictive models. However, although model [59] generated better prediction performance in terms of the AUC score, model [60] generated a 95% confidence interval. This was because the study used an imputation technique in addressing missing values and a feature selection technique to obtain an ideal MIMIC-III dataset, which resulted in a better predictive model with fine in-hospital mortality prediction performance. Using an updated version of the MIMIC-III dataset, the MIMIC-IV dataset, study [61] constructed a predictive analytics model to predict in-hospital all-cause mortality of ICU admitted patients with heart failure. The study evaluated its proposed model with the eICU-CRD database, which contains ICU-admitted heart failure-diagnosed patients. The proposed model was able to predict all-cause mortality with fine performance in terms of AUC-ROC and specificity by building predictive models using the 17 most important features selected by LASSO regression and presented in SHAP value.

3.2.3. Both Readmission and Mortality

Various studies describe the implementation of predictive analytics for the prediction of hospital readmission and mortality risk of patients with heart failure. Study [62] aimed to develop predictive analytics models to predict 30-day heart failure readmission or mortality. The study extracted a dataset from two core databases, the Hospital Morbidity Data Collection and Mortality Database of the Western Australian Data Linkage System, and used a machine learning approach to build their predictive models. By processing the data of more than 200,000 patients and applying pre-processing to the dataset, this study had good performance in predicting 30-day heart failure readmission and mortality in terms of AUC, AUPRC, and sensitivity compared to other proposed models and LACE score. Another study, [63], proposed different predictive model-building methods. By leveraging feature selection and dimension reduction techniques, this study showed improved sensitivity and the use of features lesser than their previous approach. Similar to other studies that involved a core

data system, study [64] used datasets extracted from EPIC EHR and the McKesson Change ECG Reporting System to build their predictive models for the prediction of 90-day acute heart failure readmission or all-cause mortality. The study used machine learning algorithms to build the predictive models. By assessing their proposed models with the AUC score and 95% confidence interval, the study generated predictive analytics to identify heart failure patients at high risk for either readmission or mortality, which will benefit clinicians in preventing these outcomes.

Using prospective cohort study data collected by medical centers in Shanxi Province, China, [65] generated predictive analytics models with outcomes of death, readmission, and MACEs based on patient-reported outcomes. The study evaluated AUC and Brier scores with 95% CI to each proposed predictive model based on a pre-processed dataset and SHAP value to interpret the models and to identify predictors that lead patients to hospitalization and death. For the AUC and Brier score, the proposed models were able to calculate the risk of death, readmission, and MACEs of out-of-hospital patients with heart failure. Similar to the previous study, [66] developed predictive analytics models based on electronic health records (EHR) to predict 1-year in-hospital mortality, the use of positive inotropic agents, and 1-year all-cause readmission. Their dataset was collected from the First Affiliated Hospital of Dalian Medical University, China. In building the predictive models, this study implemented machine learning algorithms and evaluated the models with their AUC-ROC, accuracy, recall, precision, and Brier Score. They then presented the SHAP values to evaluate the feature importance of the predictive models. The proposed approaches in this study showed that the EHR-driven predictive models resulted in good discrimination and predictive performances.

With GWTH-HF, an in-hospital program, study [67] derived and validated risk-prediction tools from a large nationwide registry. This study did not use machine learning algorithms in building their prediction models, but used the Center for Medicare and Medicaid Services (CMS) tool for predicting mortality and rehospitalization. Their findings claimed that the proposed models improved mortality and rehospitalization prediction. Also, their findings demonstrated fair discriminative capacity in predicting rehospitalization.

Unlike previously discussed studies, [68] and [69] used data from trial studies to develop predictive analytics for mortality risk and hospital readmission. By processing the dataset from the TOPCAT trial, which consists of 3,445 adult patients with heart failure, Study [68] found that their machine learning-based predictive models could predict the risk of mortality and rehospitalization among HFmrEF patients by the C-index and ROC-AUC score of the models. Like the previous study, [69] used the dataset from the CLEVER-HEART study, conducted between January 2008 and September 2018, to build machine learning-based predictive analytics models for predicting 30-day unplanned readmission and all-cause mortality. The assessment of the proposed models in this study yielded AUC scores of 0.723, 0.754, and 0.756, based on three different models. The proposed models performed much better than the HOSPITAL score, which had an AUC of 0.666.

4. Discussion

4.1. Machine Learning Algorithms Used for Building Predictive Models

Machine learning algorithms are essential to building and developing predictive analytics models. After data collection and pre-processing, the objective of implementing machine learning algorithms is to make the decision-making of predictive analytics more accurate in predicting specific diseases. Various studies have shown the value of implementing machine learning algorithms in their predictive models. The selected papers in this literature review presented diverse machine learning algorithms that can be used to predict heart failure, readmission, and patient mortality. As shown in Figure 2, the Random Forest algorithm appeared in 31 articles. It was useful for predicting heart failure, readmission, and mortality from the mainly used EHR-based dataset. Logistic regression, a traditional machine learning and statistics approach, also has the ability to predict heart failure in classification cases. Some studies include Logistic Regression algorithms in their model comparison because this algorithm provides a simple approach to implementation and offers an understandable interpretation.

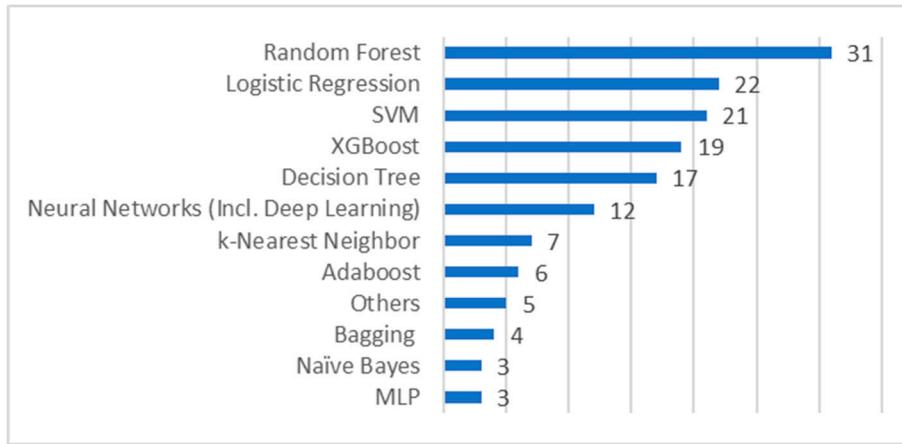


Figure 2. Machine learning algorithms used in selected research articles.

Support Vector Machine (SVM), eXtreme Gradient Boosting (XGBoost), and Decision Tree (DT) placed in the top five machine learning algorithms used for building predictive analytics for heart failure prediction. Like Random Forest, XGBoost is an ensemble of learning algorithms that uses combinations of classification trees to obtain an optimal classification result. Unlike the above ensemble learnings, SVM and Decision Trees are traditional machine learning algorithm methods that are commonly used for classification and regression. Other machine learning algorithms that seldom appeared in the elected studies are k-Nearest Neighbor, Adaboost, Bagging, Naïve Bayes, and Multilayer Perception (MLP). Other than machine learning algorithms, various studies have used neural networks and deep learning algorithms to build their predictive analytics. The use of deep learning algorithms would generate much better results than conventional machine learning approaches. However, deep learning requires enormous amounts of data, and the models tend to be challenging to interpret. Such algorithms are ubiquitous and free to use for building predictive models.

Of the studies in this literature review, five used algorithms different from those previously mentioned. Study [36] used voting feature classifier models in building their predictive analytics. [51] used classifier models modified from previously used and published approaches to building predictive analytics for the prediction of mortality for heart failure patients. By implementing a multivariable approach, [9] designed a directed acyclic graph-based model to predict heart failure from the unseen disease history of patients. Unlike other studies, [10] used the Cox Proportional Hazard model to predict heart failure. By analyzing the time to-hazard event to obtain the baseline survival probability and calculating it with the prognostic index of heart failure, the study was able to build a successful predictive model. [67] used the Center for Medicare & Medicaid Service (CMS) approach to predicting the rehospitalization and mortality of patients with heart failure. This approach is commonly used in the measurement of all-cause mortality and readmission of disease-diagnosed patients. These five studies show that various approaches and models can be implemented to build predictive analytics models that predict heart failure risk. Various measurements, scores, and scales for building mortality and readmission predictive analytics models, such as CMS, LACE, 3M PPR, and HOSPITAL, are worth considering and evaluating.

4.2. Data Pre-Processing Implementation in Building Predictive Models

In predictive analytics, obtaining an ideal dataset is a process that allows the predictive models to achieve satisfactory performance in predicting heart failure risk. As a crucial part of building a predictive model, the pre-processing step leads to much better results in predicting heart failure than can be obtained by models that do not do pre-processing. Most studies in this literature review implemented a pre-processing stage. Their results showed differences and improvements compared to studies that did not do pre-processing. Most of the studies did a simple pre-processing stage in which they reviewed the dataset manually and selected relevant features that supported the studies;

however, some studies used specific pre-processing techniques and methods that involved either algorithms or statistical approaches. By dividing pre-processing steps into data cleaning, transformation, reduction, and balancing, the following table shows the studies that used algorithms or statistical approach-based pre-processing techniques.

Table 1. Pre-processing techniques used in selected articles.

Pre-processing Step	Article	Frequent method
Data Cleaning	[7], [13], [19], [20], [23], [26], [28], [32], [33], [34], [35], [36], [39], [40], [42], [41], [45], [46], [47], [54], [55], [60], [64], [65], [66], [68], [69]	Mean imputation, predictive mean matching, median imputation, random forest imputation, kNN imputation, XGBoost imputation, and missForest
Data Transformation	[13], [14], [22], [23], [24], [25], [26], [30], [31], [34], [36], [39], [41], [42], [43], [47], [54], [55], [58], [60], [62], [33], [63], [64], [65], [69]	Recursive feature elimination, SelectKBest, Chi-Square, Pearson's correlation, KS-Test, T-Test
Data Reduction	[47], [50], [54]	PCA
Data Balancing	[13], [14], [21], [22], [23], [32], [33], [39], [43], [55], [62], [65], [66]	SMOTE, Under-sampling, Over-sampling, ADASYSN

Data cleaning includes removing missing values from the dataset and performing imputation techniques to impute the missing values. The mean imputation frequently appears in various studies, while the machine learning-based imputation technique obtained better results than statistical-based imputation. Data transformation that includes normalization, scaling, standardization, and/or feature selection for pre-processing can be used to obtain the ideal dataset. Because their data did not need to be standardized or normalized, not many of the selected studies used scaling, normalization, or standardization. However, feature selection appears in many studies in this literature review. The most used feature selection technique is Chi-square, and in various studies, this feature selection technique gives a promising performance result. Many studies used machine learning-based feature selection in post-processing to describe and present the most important features that influenced their prediction. Data reduction in the above table shows the use of dimensionality reduction to produce small dimensions of features in a dataset. Principal component analysis, or PCA, was the most frequently used type in the selected articles. Data balancing was done in studies with imbalanced classes to obtain the ideal dataset. The most used data balancing technique was the SMOTE method. Some selected papers mentioned under-sampling and over-sampling methods, but they did not mention the specific names of the methods. Implementing a pre-processing technique would remarkably improve prediction performance. Pre-processing techniques like feature selection can also be used in post-processing to present the most important features that would allow clinicians and experts to understand the predictive models' results in predicting heart failure risks. Many pre-processing techniques can be used in the development of predictive analytics; however, critical evaluation of each technique is necessary to generate the most ideal predictive models.

4.3. Data Specification Used in Building Predictive Models

The dataset is a critical factor in building heart failure risk predictive analytics models. However, to find an ideal dataset we need to first understand the dataset specifications necessary for building models that can predict the target disease. Demographic data such as age, gender, and race frequently appear in published studies and are used in their predictive analytics building. Study [47], which involved experts in selecting the features most related to heart failure risk prediction, reported that gender, age, and body mass index; expert-recommended demographical data; were the features most closely related to the risk of heart failure. For readmission prediction, [62] and [63] used a t-test approach and found age to be the leading feature for predicting 30-day heart failure readmission or death. These two studies give evidence of the importance of the use of demographical data. Through

the use of an open dataset from the UCI database with SelectKBest and the chi-square method, study [23] showed that sex and age impacted heart failure prediction.

The results for the predictive analytics models that predicted heart failure and readmission did not result in differences in their demographic data, with the various studies finding similar correlations. Age and sex appeared 40 and 37 times in the papers for heart failure and remission, respectively. Similar to demographic data, the physical and clinical values of patients were essential to building an effective predictive analytics model for heart failure prediction. Physical and clinical values such as weight, diastolic and systolic blood pressure, and heart rate are patient data that clinicians routinely collect. According to the selected papers, both diastolic and systolic blood pressure, heart rate, weight or BMI, respiratory rate, and temperature, appeared 22, 19, 18, 10, and 7 times, respectively. This indicates the importance of physical and clinical values in improving the prediction performance of heart failure predictive analytics models. Because these values have been shown to be indicators of heart conditions, they are essential for building an accurate heart failure predictive model. Other variables, such as lifestyle variables should also be considered.

In the selected papers, smoking and alcohol appeared 15 and 6 times, respectively. Both would be useful for heart failure risk prediction because they have been shown to be related to heart disease. In various studies, comorbidities have been essential to building predictive models for the prediction of heart failure or readmission, appearing 23 times in the selected papers. Study [52] showed that including comorbidity data would lead to prediction performance improvement compared to models without comorbidity data. The comorbidities of patients might be related to the worsening of their condition, especially for patients with heart failure. The figure below summarizes the data categories described in the selected papers.

Note that various papers used the same open data, so we combined them. For example, ten papers used the UCI heart disease dataset, so we reported their data together in the table. The number presented in Figure 3 is the total times specific data types appeared in the selected papers. According to the attached figure, various studies used data like medication information and ECG values in building their models. However, a few studies used unstructured data like doctor notes. Study [43] used feature selection from unstructured data to predict hospital readmission. By processing patient data, such as patient descriptions, clinical notes, discharge summaries, diagnoses and procedures, and admission events, the study was able to generate models that predicted readmission of heart failure patients, with good AUC performance scores. Processing unstructured data to build a heart failure predictive model is suitable for in-house applications like telemedicine, where the clinician collects patient data by phone. However, in most cases additional data like demographic data and physical and clinical values are necessary to improve the prediction. Understanding the dataset is essential to building the ideal and optimal predictive analytics models for heart failure prediction. We can generate predictive models with satisfactory prediction performance by selecting the right features and variables and choosing a suitable model.

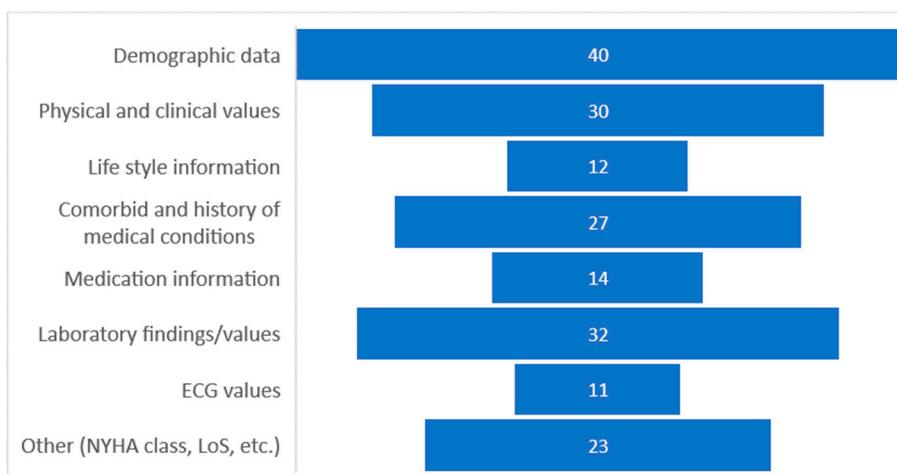


Figure 3. Data categories reported in the selected papers.

4.4. Publication by Year

Figure 4 below presents the publication trend of papers reporting predictive analytics for heart failure prediction based on refined search results. The graph clearly shows a yearly increase in published papers using predictive analytics. The highest increase was seen in 2021, with about 130 papers published.

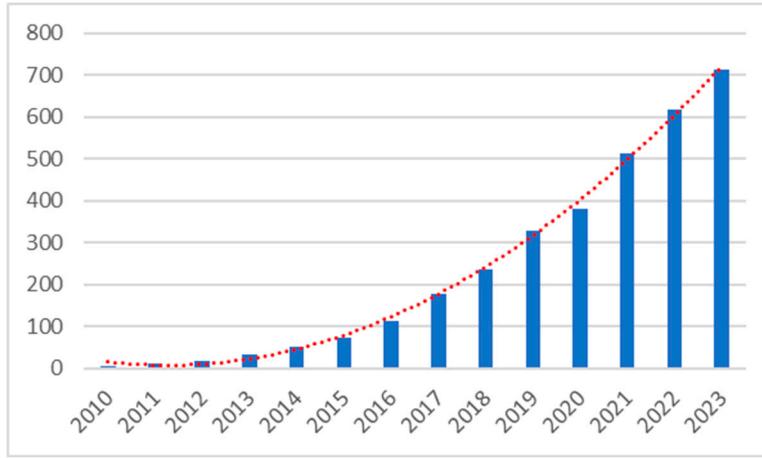


Figure 4. Publication by year based on refined-search results.

We broke down the selected papers above into four well-known methodologies used in the building of predictive models, as in Figure 5. The first is a medical approach, commonly used by medical field experts like doctors and clinicians to develop predictive models based on medical or clinical approaches such as physical examination and laboratory tests. The second is a statistical approach to modeling the predictive analytics. There are various statistical analysis approaches, including the Cox hazard proportional approach. The third is machine learning, an artificial intelligence-based approach that is widespread among researchers and that is especially useful for developing predictive analytics models. The last methodology uses big data and data mining through which researchers use various big data applications to build predictive models. The researchers of a few studies combined big data and machine learning approaches.

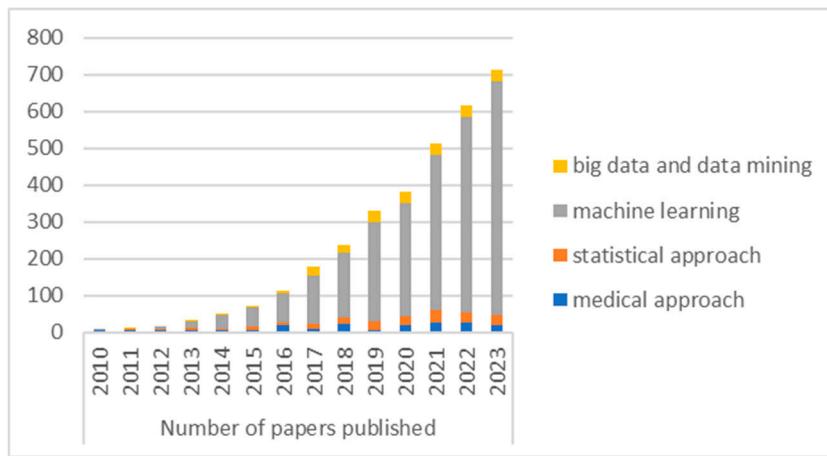


Figure 5. Methodology used in heart failure predictive analytics papers between 2010 and 2023.

5. Conclusions

Heart failure is one of the most common and deadliest diseases among people worldwide, and many clinicians and experts are hoping for the development of useful methods of predicting the risk of heart failure. The use of technologies such as machine learning makes the predicting heart failure risk much easier and more accurate and is of great value to patients. Predictive analytics can be used

by clinicians and other experts to predict heart failure and readmission. Many studies have been done worldwide in the last decade to build heart failure predictive analytics models. Future predictive analytics models with more accurate and timely prediction performance will be important to the decision-making of clinicians and other experts.

Author Contributions: Conceptualization, Q.H. and E.H.; methodology, Q.H.; original draft preparation, Q.H.; review and editing, Q.H. and E.H.; supervision, E.H.

Funding: There is no funding for this study.

Institutional Review Board Statement: There is no Institutional Review Board statement required for this review paper.

Informed Consent Statement: Not applicable.

Data Availability Statement: As it is a review paper, all data we used in this study can be available from public databases such as PubMed and Google Scholar.

Acknowledgments: Authors would like to thank the members of the Laboratory of Fundamental and Applied Informatics who supported this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Association, A.H. What Is Heart Failure?. Available online: <https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure>
2. Ponikowski, P.; Anker, S.D.; AlHabib, K.F.; Cowie, M.R.; Force, T.L.; Hu, S.; Jaarsma, T.; Krum, H.; Rastogi, V.; Rohde, L.E.; et al. Heart Failure: Preventing Disease and Death Worldwide. *ESC Heart Failure* 2014, 1, 4–25, doi:10.1002/ehf2.12005.
3. Jencks, S.F.; Williams, M.V.; Coleman, E.A. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. *N Engl J Med* 2009, 360, 1418–1428, doi:10.1056/NEJMsa0803563.
4. Chen, M.; Hao, Y.; Hwang, K.; Wang, L.; Wang, L. Disease Prediction by Machine Learning Over Big Data From Healthcare Communities. *IEEE Access* 2017, 5, 8869–8879, doi:10.1109/ACCESS.2017.2694446.
5. Guidi, G.; Pettenati, M.C.; Melillo, P.; Iadanza, E. A Machine Learning System to Improve Heart Failure Patient Assistance. *IEEE J. Biomed. Health Inform.* 2014, 18, 1750–1756, doi:10.1109/JBHI.2014.2337752.
6. Yajuan Wang; Ng, K.; Byrd, R.J.; Jianying Hu; Ebadollahi, S.; Daar, Z.; deFilippi, C.; Steinhubl, S.R.; Stewart, W.F. Early Detection of Heart Failure with Varying Prediction Windows by Structured and Unstructured Data in Electronic Health Records. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); IEEE: Milan, August 2015; pp. 2530–2533, doi:10.1109/EMBC.2015.7318907.
7. Ng, K.; Steinhubl, S.R.; deFilippi, C.; Dey, S.; Stewart, W.F. Early Detection of Heart Failure Using Electronic Health Records: Practical Implications for Time Before Diagnosis, Data Diversity, Data Quantity, and Data Density. *Circ: Cardiovascular Quality and Outcomes* 2016, 9, 649–658, doi:10.1161/CIRCOUTCOMES.116.002797.
8. Rammal, H.F.; Z., A. Heart Failure Prediction Models Using Big Data Techniques. *ijacsa* 2018, 9, doi:10.14569/IJACSA.2018.090547.
9. Nagrecha, S.; Thomas, P.B.; Feldman, K.; Chawla, N.V. Predicting Chronic Heart Failure Using Diagnoses Graphs. In Machine Learning and Knowledge Extraction; Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, 2017; Vol. 10410, pp. 295–312, doi:10.1007/978-3-319-66808-6_20.
10. Krittayaphong, R.; Chichareon, P.; Komoltri, C.; Sairat, P.; Lip, G.Y.H. Predicting Heart Failure in Patients with Atrial Fibrillation: A Report from the Prospective COOL-AF Registry. *Journal of clinical medicine* 2023, 12, 4, 1265, doi: 10.3390/jcm12041265.
11. Austin, P.C.; Tu, J.V.; Ho, J.E.; Levy, D.; Lee, D.S. Using Methods from the Data-Mining and Machine-Learning Literature for Disease Classification and Prediction: A Case Study Examining Classification of Heart Failure Subtypes. *Journal of Clinical Epidemiology* 2013, 66, 398–407, doi:10.1016/j.jclinepi.2012.11.008.

12. Blecker, S.; Sontag, D.; Horwitz, L.I.; Kuperman, G.; Park, H.; Reyentovich, A.; Katz, S.D. Early Identification of Patients With Acute Decompensated Heart Failure. *Journal of Cardiac Failure* 2018, 24, 357–362, doi:10.1016/j.cardfail.2017.08.458.
13. Plati, D.K.; Tripoliti, E.E.; Bechlioulis, A.; Rammos, A.; Dimou, I.; Lakkas, L.; Watson, C.; McDonald, K.; Ledwidge, M.; Pharithi, R.; et al. A Machine Learning Approach for Chronic Heart Failure Diagnosis. *Diagnostics* 2021, 11, 1863, doi:10.3390/diagnostics11101863.
14. Wang, K.; Tian, J.; Zheng, C.; Yang, H.; Ren, J.; Li, C.; Han, Q.; Zhang, Y. Improving Risk Identification of Adverse Outcomes in Chronic Heart Failure Using SMOTE+ENN and Machine Learning. *RMHP* 2021, Volume 14, 2453–2463, doi:10.2147/RMHP.S310295.
15. Quesada, J.A.; Lopez-Pineda, A.; Gil-Guillén, V.F.; Durazo-Arvizu, R.; Orozco-Beltrán, D.; López-Domenech, A.; Carratalá-Munuera, C. Machine Learning to Predict Cardiovascular Risk. *Int J Clin Pract* 2019, 73, doi:10.1111/ijcp.13389.
16. Kolukula, N.R.; Pothineni, P.N.; Chinta, V.M.K.; Boppana, V.G.; Kalapala, R.P.; Duvvi, S. Predictive Analytics of Heart Disease Presence with Feature Importance Based on Machine Learning Algorithms. *IJEECS* 2023, 32, 1070, doi:10.11591/ijeecs.v32.i2.pp1070-1077.
17. Sornsuwit, P.; Jundahuadong, P.; Pongsakornrungsilp, S. A New Efficiency Improvement of Ensemble Learning for Heart Failure Classification by Least Error Boosting. *Emerg Sci J* 2022, 7, 135–146, doi:10.28991/ESJ-2023-07-01-010.
18. Ahmed, S.; Shaikh, S.; Ikram, F.; Fayaz, M.; Alwageed, H.S.; Khan, F.; Jaskani, F.H. Prediction of Cardiovascular Disease on Self-Augmented Datasets of Heart Patients Using Multiple Machine Learning Models. *Journal of Sensors* 2022, 2022, 1–21, doi:10.1155/2022/3730303.
19. Praveena Rachel Kamala, S.; Gayathri, S.; Pillai, N.M.; Anto Gracious, L.A.; Varun, C.M.; Siva Subramanian, R. Predictive Analytics for Heart Disease Detection: A Machine Learning Approach. In Proceedings of the 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC); IEEE: Coimbatore, India, July 6 2023; pp. 1583–1589, doi:10.1109/ICESC57686.2023.10193323.
20. Alotaibi, F.S. Implementation of Machine Learning Model to Predict Heart Failure Disease. *IJACSA* 2019, 10, doi:10.14569/IJACSA.2019.0100637.
21. Mamun, M.; Farjana, A.; Mamun, M.A.; Ahammed, M.S.; Rahman, M.M. Heart Failure Survival Prediction Using Machine Learning Algorithm: Am I Safe from Heart Failure? In Proceedings of the 2022 IEEE World AI IoT Congress (AIIoT); IEEE: Seattle, WA, USA, June 6 2022; pp. 194–200, doi:10.1109/AIIoT54504.2022.9817303.
22. Nishat, M.M.; Faisal, F.; Ratul, I.J.; Al-Monsur, A.; Ar-Rafi, A.M.; Nasrullah, S.M.; Reza, T.; Khan, R.H. A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset. *Scientific Programming*, 2022, 1–17, doi:10.1155/2022/3649406.
23. Senan, E.M.; Abunadi, I.; Jadhav, M.E.; Fati, S.M. Score and Correlation Coefficient-Based Feature Selection for Predicting Heart Failure Diagnosis by Using Machine Learning Algorithms. *Computational and Mathematical Methods in Medicine* 2021, 2021, 1–16, doi:10.1155/2021/8500314.
24. Al-Yarimi, F.A.M.; Munassar, N.M.A.; Bamashmos, M.H.M.; Ali, M.Y.S. Feature Optimization by Discrete Weights for Heart Disease Prediction Using Supervised Learning. *Soft Comput* 2021, 25, 1821–1831, doi:10.1007/s00500-020-05253-4.
25. Bharti, R.; Khamparia, A.; Shabaz, M.; Dhiman, G.; Pande, S.; Singh, P. Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. *Computational Intelligence and Neuroscience* 2021, 2021, 1–11, doi:10.1155/2021/8387680.
26. Kanagarathinam, K.; Sankaran, D.; Manikandan, R. Machine Learning-Based Risk Prediction Model for Cardiovascular Disease Using a Hybrid Dataset. *Data & Knowledge Engineering* 2022, 140, 102042, doi:10.1016/j.dake.2022.102042.
27. Venkatesh, R.; Balasubramanian, C.; Kaliappan, M. Development of Big Data Predictive Analytics Model for Disease Prediction Using Machine Learning Technique. *J Med Syst* 2019, 43, 272, doi:10.1007/s10916-019-1398-y.
28. Alsubai, S.; Alqahtani, A.; Binbusayyis, A.; Sha, M.; Gumei, A.; Wang, S. Heart Failure Detection Using Instance Quantum Circuit Approach and Traditional Predictive Analysis. *Mathematics* 2023, 11, 1467, doi:10.3390/math11061467.

29. Botros, J.; Mourad-Chehade, F.; Laplanche, D. CNN and SVM-Based Models for the Detection of Heart Failure Using Electrocardiogram Signals. *Sensors* 2022, 22, 9190, doi:10.3390/s22239190.
30. Alsinglawi, B.; Alnajjar, F.; Mubin, O.; Novoa, M.; Alorjani, M.; Karajeh, O.; Darwish, O. Predicting Length of Stay for Cardiovascular Hospitalizations in the Intensive Care Unit: Machine Learning Approach. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); IEEE: Montreal, QC, Canada, July 2020; pp. 5442–5445, doi:10.1109/EMBC44109.2020.9175889.
31. Shameer, K.; Johnson, K.W.; Yahi, A.; Miotto, R.; Li, L.; Ricks, D.; Jebakaran, J.; Kovatch, P.; Sengupta, P.P.; Gelijns, S.; et al. PREDICTIVE MODELING OF HOSPITAL READMISSION RATES USING ELECTRONIC MEDICAL RECORD-WIDE MACHINE LEARNING: A CASE-STUDY USING MOUNT SINAI HEART FAILURE COHORT. In Proceedings of the Biocomputing 2017; WORLD SCIENTIFIC: Kohala Coast, Hawaii, USA, January 2017; pp. 276–287, doi:10.1142/9789813207813_0027.
32. Bat-Erdene, B.-I.; Zheng, H.; Son, S.H.; Lee, J.Y. Deep Learning-Based Prediction of Heart Failure Rehospitalization during 6, 12, 24-Month Follow-Ups in Patients with Acute Myocardial Infarction. *Health Informatics J* 2022, 28, 146045822211015, doi:10.1177/14604582221101529.
33. Rizinde, T.; Ngaruye, I.; Cahill, N.D. Comparing Machine Learning Classifiers for Predicting Hospital Readmission of Heart Failure Patients in Rwanda. *JPM* 2023, 13, 1393, doi:10.3390/jpm13091393.
34. Landicho, J.A.; Esichaikul, V.; Sasil, R.M. Comparison of Predictive Models for Hospital Readmission of Heart Failure Patients with Cost-Sensitive Approach. *International Journal of Healthcare Management* 2021, 14, 1536–1541, doi:10.1080/20479700.2020.1797334.
35. Sohrabi, B.; Vanani, I.R.; Gooyavar, A.; Naderi, N. Predicting the Readmission of Heart Failure Patients through Data Analytics. *J. Info. Know. Mgmt.* 2019, 18, 1950012, doi:10.1142/S0219649219500126.
36. AbdelRahman, S.E.; Zhang, M.; Bray, B.E.; Kawamoto, K. A Three-Step Approach for the Derivation and Validation of High-Performing Predictive Models Using an Operational Dataset: Congestive Heart Failure Readmission Case Study. *BMC Med Inform Decis Mak* 2014, 14, 41, doi:10.1186/1472-6947-14-41.
37. Vedomske, M.A.; Brown, D.E.; Harrison, J.H. Random Forests on Ubiquitous Data for Heart Failure 30-Day Readmissions Prediction. In Proceedings of the 2013 12th International Conference on Machine Learning and Applications; IEEE: Miami, FL, December 2013; pp. 415–421, doi:10.1109/ICMLA.2013.158.
38. Hilbert, J.P.; Zasadil, S.; Keyser, D.J.; Peele, P.B. Using Decision Trees to Manage Hospital Readmission Risk for Acute Myocardial Infarction, Heart Failure, and Pneumonia. *Appl Health Econ Health Policy* 2014, 12, 573–585, doi:10.1007/s40258-014-0124-7.
39. Zolbanin, H.M.; Delen, D. Processing Electronic Medical Records to Improve Predictive Analytics Outcomes for Hospital Readmissions. *Decision Support Systems* 2018, 112, 98–110, doi:10.1016/j.dss.2018.06.010.
40. Golas, S.B.; Shibahara, T.; Agboola, S.; Otaki, H.; Sato, J.; Nakae, T.; Hisamitsu, T.; Kojima, G.; Felsted, J.; Kakarmath, S.; et al. A Machine Learning Model to Predict the Risk of 30-Day Readmissions in Patients with Heart Failure: A Retrospective Analysis of Electronic Medical Records Data. *BMC Med Inform Decis Mak* 2018, 18, 44, doi:10.1186/s12911-018-0620-z.
41. Mortazavi, B.J.; Downing, N.S.; Bucholz, E.M.; Dharmarajan, K.; Manhapra, A.; Li, S.-X.; Negahban, S.N.; Krumholz, H.M. Analysis of Machine Learning Techniques for Heart Failure Readmissions. *Circ: Cardiovascular Quality and Outcomes* 2016, 9, 629–640, doi:10.1161/CIRCOUTCOMES.116.003039.
42. Lorenzoni, G.; Sabato, S.S.; Lanera, C.; Bottigliengo, D.; Minto, C.; Ocagli, H.; De Paolis, P.; Gregori, D.; Iliceto, S.; Pisanò, F. Comparison of Machine Learning Techniques for Prediction of Hospitalization in Heart Failure Patients. *JCM* 2019, 8, 1298, doi:10.3390/jcm8091298.
43. Sundararaman, A.; Valady Ramanathan, S.; Thati, R. Novel Approach to Predict Hospital Readmissions Using Feature Selection from Unstructured Data with Class Imbalance. *Big Data Research* 2018, 13, 65–75, doi:10.1016/j.bdr.2018.05.004.
44. Liu, X.; Chen, Y.; Bae, J.; Li, H.; Johnston, J.; Sanger, T. Predicting Heart Failure Readmission from Clinical Notes Using Deep Learning. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); IEEE: San Diego, CA, USA, November 2019; pp. 2642–2648, doi:10.1109/BIBM47256.2019.8983095.
45. Sharma, V.; Kulkarni, V.; McAlister, F.; Eurich, D.; Keshwani, S.; Simpson, S. H.; Voaklander, D.; Samanani, S. Predicting 30-Day Readmissions in Patients With Heart Failure Using Administrative Data: A Machine Learning Approach. *Journal of Cardiac Failure* 2022, 28, 5, 710–722, doi:10.1016/j.cardfail.2021.12.004.

46. Shams, I.; Ajorlou, S.; Yang, K. A Predictive Analytics Approach to Reducing 30-Day Avoidable Readmissions among Patients with Heart Failure, Acute Myocardial Infarction, Pneumonia, or COPD. *Health Care Manag Sci* 2015, 18, 19–34, doi:10.1007/s10729-014-9278-y.

47. Ben-Assuli, O.; Heart, T.; Klempfner, R.; Padman, R. Human-Machine Collaboration for Feature Selection and Integration to Improve Congestive Heart Failure Risk Prediction. *Decision Support Systems* 2023, 172, 113982, doi:10.1016/j.dss.2023.113982.

48. Jing, L.; Ulloa Cerna, A.E.; Good, C.W.; Sauers, N.M.; Schneider, G.; Hartzel, D.N.; Leader, J.B.; Kirchner, H.L.; Hu, Y.; Riviello, D.M.; et al. A Machine Learning Approach to Management of Heart Failure Populations. *JACC: Heart Failure* 2020, 8, 578–587, doi:10.1016/j.jchf.2020.01.012.

49. Kamio, T.; Ikegami, M.; Machida, Y.; Uemura, T.; Chino, N.; Iwagami, M. Machine Learning-Based Prognostic Modeling of Patients with Acute Heart Failure Receiving Furosemide in Intensive Care Units. *DIGITAL HEALTH* 2023, 9, 20552076231194933, doi:10.1177/20552076231194933.

50. Adler, E.D.; Voors, A.A.; Klein, L.; Macheret, F.; Braun, O.O.; Urey, M.A.; Zhu, W.; Sama, I.; Tadel, M.; Campagnari, C.; et al. Improving Risk Prediction in Heart Failure Using Machine Learning. *European J of Heart Fail* 2020, 22, 139–147, doi:10.1002/ejhf.1628.

51. Lagu, T.; Pekow, P.S.; Shieh, M.-S.; Stefan, M.; Pack, Q.R.; Kashef, M.A.; Atreya, A.R.; Valania, G.; Slawsky, M.T.; Lindenauer, P.K. Validation and Comparison of Seven Mortality Prediction Models for Hospitalized Patients With Acute Decompensated Heart Failure. *Circ: Heart Failure* 2016, 9, e002912, doi:10.1161/CIRCHEARTFAILURE.115.002912.

52. Panahiazar, M.; Taslimitehrani, V.; Pereira, N.; Pathak, J. Using EHRs and Machine Learning for Heart Failure Survival Analysis. *Studies in health technology* 2015, 2016, 40–44, doi:10.3233/978-1-61499-564-7-40.

53. Almazroi, A.A. Survival Prediction among Heart Patients Using Machine Learning Techniques. *MBE* 2022, 19, 134–145, doi:10.3934/mbe.2022007.

54. Özbay Karakuş, M.; Er, O. A Comparative Study on Prediction of Survival Event of Heart Failure Patients Using Machine Learning Algorithms. *Neural Comput & Applic* 2022, 34, 13895–13908, doi:10.1007/s00521-022-07201-9.

55. Zaman, S.M.M.; Qureshi, W.M.; Raihan, M.S.; Shams, A.B.; Sultana, S. Survival Prediction of Heart Failure Patients Using Stacked Ensemble Machine Learning Algorithm. In 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering. IEEE: Dhaka, Bangladesh, 19 July 2022; pp. 117–120, doi:10.1109/WIECON-ECE54711.2021.9829577.

56. Newaz, A.; Ahmed, N.; Shahriyar Haq, F. Survival Prediction of Heart Failure Patients Using Machine Learning Techniques. *Informatics in Medicine Unlocked* 2021, 26, 100772, doi:10.1016/j.imu.2021.100772.

57. Kedia, S.; Bhushan, M. Prediction of Mortality from Heart Failure Using Machine Learning. In Proceedings of the 2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET); IEEE: Patna, India, June 24 2022; pp. 1–6, doi:10.1109/ICEFEET51821.2022.9848348.

58. Chicco, D.; Jurman, G. Machine Learning Can Predict Survival of Patients with Heart Failure from Serum Creatinine and Ejection Fraction Alone. *BMC Med Inform Decis Mak* 2020, 20, 16, doi:10.1186/s12911-020-1023-5.

59. Li, F.; Xin, H.; Zhang, J.; Fu, M.; Zhou, J.; Lian, Z. Prediction Model of In-hospital Mortality in Intensive Care Unit Patients with Heart Failure: Machine Learning-Based, Retrospective Analysis of the MIMIC-III Database. *BMJ open* 2021, 11, 7, doi:10.1136/bmjopen-2020-044779.

60. Luo, C.; Zhu, Y.; Zhu, Z.; Li, R.; Chen, G.; Wang, Z. A Machine Learning-Based Risk Stratification Tool for in-Hospital Mortality of Intensive Care Unit Patients with Heart Failure. *J Transl Med* 2022, 20, 136, doi:10.1186/s12967-022-03340-8.

61. Chen, Z.; Li, T.; Guo, S.; Zeng, D.; Wang, K. Machine Learning-Based in-Hospital Mortality Risk Prediction Tool for Intensive Care Unit Patients with Heart Failure. *Frontiers in Cardiovascular Medicine* 2023, 10, 1119699, doi:10.3389/fcvm.2023.1119699.

62. Awan, S.E.; Bennamoun, M.; Sohel, F.; Sanfilippo, F.M.; Dwivedi, G. Machine Learning-based Prediction of Heart Failure Readmission or Death: Implications of Choosing the Right Model and the Right Metrics. *ESC Heart Failure* 2019, 6, 428–435, doi:10.1002/ehf2.12419.

63. Awan, S.E.; Bennamoun, M.; Sohel, F.; Sanfilippo, F.M.; Chow, B.J.; Dwivedi, G. Feature Selection and Transformation by Machine Learning Reduce Variable Numbers and Improve Prediction for Heart Failure Readmission or Death. *PLoS ONE* 2019, 14, e0218760, doi:10.1371/journal.pone.0218760.

64. Sarjaloo, F.; Park, J.; Zhong, X.; Wokhlu, A. Predicting 90 DAY Acute Heart Failure Readmission and Death Using Machine LEARNING-SUPPORTED Decision Analysis. *Clinical Cardiology* 2021, **44**, 230–237, doi:10.1002/clc.23532.
65. Tian, J.; Yan, J.; Han, G.; Du, Y.; Hu, X.; He, Z.; Han, Q.; Zhang, Y. Machine Learning Prognosis Model Based on Patient-Reported Outcomes for Chronic Heart Failure Patients after Discharge. *Health Qual Life Outcomes* 2023, **21**, 31, doi:10.1186/s12955-023-02109-x.
66. Lv, H.; Yang, X.; Wang, B.; Wang, S.; Du, X.; Tan, Q.; Hao, Z.; Liu, Y.; Yan, J.; Xia, Y. Machine Learning–Driven Models to Predict Prognostic Outcomes in Patients Hospitalized With Heart Failure Using Electronic Health Records: Retrospective Study. *J Med Internet Res* 2021, **23**, e24996, doi:10.2196/24996.
67. Eapen, Z.J.; Liang, L.; Fonarow, G.C.; Heidenreich, P.A.; Curtis, L.H.; Peterson, E.D.; Hernandez, A.F. Validated, Electronic Health Record Deployable Prediction Models for Assessing Patient Risk of 30-Day Rehospitalization and Mortality in Older Heart Failure Patients. *JACC: Heart Failure* 2013, **1**, 245–251, doi:10.1016/j.jchf.2013.01.008.
68. Zhao, H.; Li, P.; Zhong, G.; Xie, K.; Zhou, H.; Ning, Y.; Xu, D.; Zeng, Q. Machine Learning Models in Heart Failure with Mildly Reduced Ejection Fraction Patients. *Front. Cardiovasc. Med.* 2022, **9**, 1042139, doi:10.3389/fcvm.2022.1042139.
69. Beecy, A.N.; Gummalla, M.; Sholle, E.; Xu, Z.; Zhang, Y.; Michalak, K.; Dolan, K.; Hussain, Y.; Lee, B.C.; Zhang, Y.; et al. Utilizing Electronic Health Data and Machine Learning for the Prediction of 30-Day Unplanned Readmission or All-Cause Mortality in Heart Failure. *Cardiovascular Digital Health Journal* 2020, **1**, 71–79, doi:10.1016/j.cvdhj.2020.07.004.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.