Article

Comparison of Selected Machine Learning Algorithms in the Analysis of Mental Health Indicators

Adrian Bieliński, Izabela Rojek * and Dariusz Mikołajewski

Faculty of Computer Science, Kazimierz Wielki University, 85-064 Bydgoszcz, Poland;
adrian.bielinski.pl@gmail.com (A.B.); dariusz.mikolajewski@ukw.edu.pl (D.M.)
* Correspondence: izabela.rojek@ukw.edu.pl

Abstract: Machine learning is increasingly being used to solve clinical problems in diagnosis, therapy and care. Aim: the main aim of the study was to investigate how the selected machine learning algorithms deal with the problem of determining a virtual mental health index. Material and Methods: a number of machine learning models based on Stochastic Dual Coordinate Ascent, limited-memory Broyden–Fletcher–Goldfarb–Shanno, Online Gradient Descent, etc., were built based on a clinical dataset and compared based on criteria in the form of learning time, running time during use and regression accuracy. Results: the algorithm with the highest accuracy was Stochastic Dual Coordinate Ascent, but although its performance was high, it had significantly longer training and prediction times. The fastest algorithm looking at learning and prediction time, but slightly less accurate, was the limited-memory Broyden–Fletcher–Goldfarb–Shanno. The same data set was also analyzed automatically using ML.NET. Findings from the study can be used to build larger systems that automate early mental health diagnosis and help differentiate the use of individual algorithms depending on the purpose of the system.

Keywords: computer science; artificial intelligence; machine learning; burnout; clinical reasoning

1. Introduction

The development of machine learning (ML) is driven by the vast amount of data available (so-called big data), which are used to train algorithms to adapt them to solve scientific, clinical and industrial problems quickly and efficiently [1,2]. ML is a data-driven approach in which rules are extracted automatically based on associations between input and output data sets, and their relevance is tested against validation data. Models learned in this way (mainly traditional and deep artificial neural networks) can then be trained to better fit new data. Machine learning is increasingly being used to solve clinical problems in diagnosis, therapy and care [3–5]. The number of publications on clinical applications of machine learning increased rapidly after 2010, with the main areas of research being in diagnostics and prediction, and less often in classical clinical problem solving (Figure 1a–d).

In recent years, there has been a growing interest in the application of ML in the diagnosis (less frequently: therapy) of mental health (Figure 1e) [6,7]. This is due to a number of factors, but above all to the fact that this group of conditions is becoming common as a new group of diseases of civilization in adults, children and adolescents, while at the same time representing very complex and stigmatizing disease entities that are difficult to combat with limited resources and numbers of specialists. Automation of certain procedures is therefore possible and desirable for both patients and medical staff.
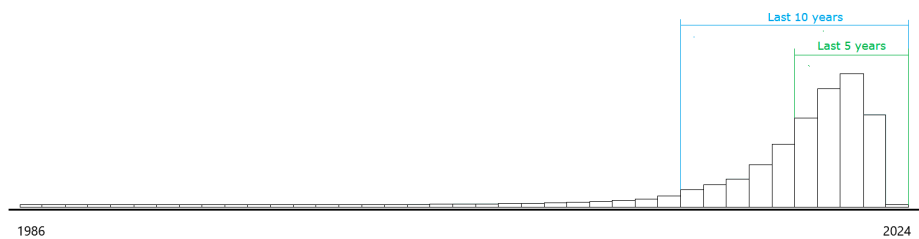
The main aim of the study was to see how the selected ML algorithms deal with the problem of determining a virtual mental health index.

(**a**)

1964   2024

(**b**)

1994   2023

(**c**)

1986   2024

(**d**)

1988   2024

(**e**)

2009   2023
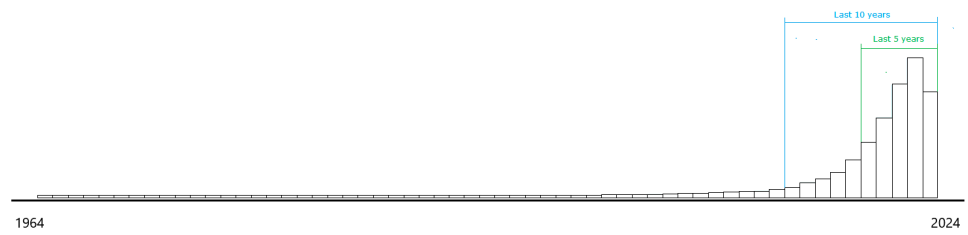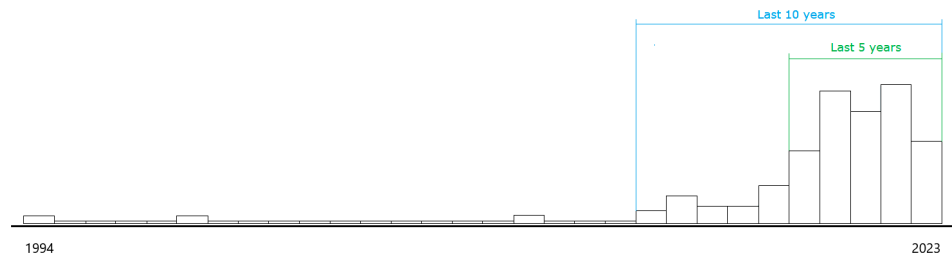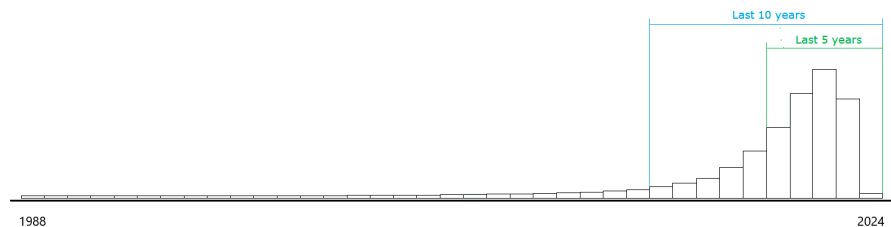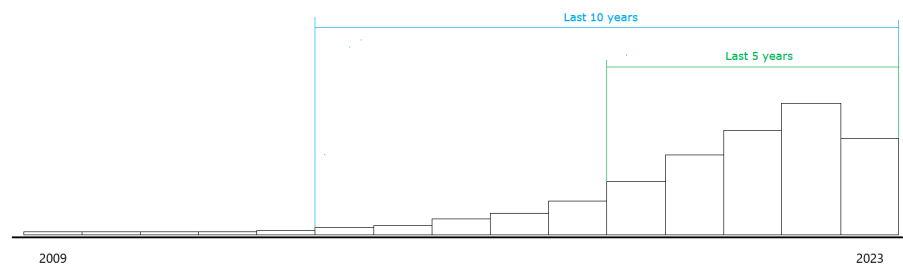
Figure 1. Number of scientific publications: (**a**) concerning clinical applications of machine learning (total number of publications: 103,017), (**b**)with keywords "machine learning" and "clinical problem solving" (total number of publications: 113), (**c**) with keywords "machine learning" and "diagnosis" (total number of publications: 37,242), (**d**) with keywords "machine learning" and "prediction" (total number of publications: 50,619), and (**e**) with keywords "machine learning" and "mental health" (2332).

Related Publications

There are many articles in the literature on the virtual mental health index. Each of them stands out from the others, approaching the topic from a different point of view. One article addresses the topic of e-health and modern technologies used in mental health care [8,9]. It is indicated that the aim of the article is to present issues related to e-health, and its elements used in the diagnosis and treatment of patients with mental disorders. The article points out that there is a lot of enthusiasm for e-health issues around the world, which may be related to the transformation potential of the healthcare system [8,9]. The article points out that e-health solutions have been shown to be effective in preventing, diagnosing and treating patients with a variety of illnesses, both physical and mental [9], including substance abuse, depression, bipolar disorder, anxiety, stress and/or suicidal thoughts. This article adopts the World Health Organisation's (WHO) definition of e-health. In addition, differences between the original and the newer definition are pointed out, as the newer definition describes it as the use of electronic means of communicating health-related information, resources and services, whereas the original definition presented the concept as the use of information technology, locally and remotely, in support of health and related fields. The newer definition according to the WHO also includes electronic health records, mobile health and health analytics. An important change was also indicated in the context of the patient–professional relationship, i.e., the patient participates as a partner in the diagnosis and treatment process, rather than being merely a passive figure. An increase in patients' responsibility for their own treatment, an increase in their involvement in treatment decisions or a tendency to use strengthening and improvement exercises were also noted. It was also mentioned that inviting the patient into the e-health system does not imply patient involvement. The studies mentioned in this article identified three different types of involvement: active, partner and submissive [8,9]. Mobile apps used in practice were also identified, including for practicing stress management skills, in the diagnosis and treatment of depression, and as an aid to screening. The cited authors indicated that apps could be used to monitor mental status and mood, as well as bipolar affective disorder [8,9]. This article presents modern technology as an opportunity for the development of medicine, including in the context of mental health. The article draws on a number of sources, indicating that these are not isolated, exceptional situations. It is noteworthy that it was written before the onset of the problems associated with the COVID-19 pandemic. This article provides an interesting insight into the applications of technology not only in treatment but also in prevention. In contrast, another article [10] deals with the use of ML techniques to predict stress in active workers. As an introduction, the prevalence of mental disorders among the working class was highlighted, with a clear upward trend when looking at the percentage of employees who experience depressive and anxious states. It was concluded that the greatest emphasis must be placed on maintaining a stress-free atmosphere in order to achieve better productivity and well-being of employees. The authors [10] used the results of a survey of technology employees in 2017, with which they trained various models for their analyses. The original data consisted of 750 responses from people from different technical departments in the form of 68 attributes related to private life and work. A data cleaning exercise was carried out, which left 14 parameters, in addition to which a one-hot encoding (1 of n) was used to represent some fields as numeric. In addition, the text responses 'Yes' were given a value of 1, 'No' a value of 0, and 'Maybe' a value of 0.5. NaN values were replaced by 0, and nominal data were converted to numeric using a label encoder. The authors chose models for training that had already been tested in classification problems, implementing them in Python using the Scikit-learn library:

- Logistic regression;
- K-nearest-neighbormethod;
- Decision trees;
- Random forest;
- Boosting (increasing the effectiveness of existing models);
- Bagging.

The following were used as metrics of model performance:

- Classification accuracy;
- False Positive Rate, which indicates how many negative cases were classified as positive;
- Precision, i.e., the fraction of cases predicted to be positive that were actually positive;
- Area Under the Curve (AUC) score;
- Cross-validation AUC score [10].

Each model assessed whether a person required treatment. These tests resulted in model accuracies ranging from 69.43% to 75.13%, with the bagging algorithm achieving the lowest level and the boosting algorithm the highest. The greatest influence on stress and mental health was the gender of the individual, as well as family history and the services provided by the employing entity for mental health care. As further research opportunities, the authors [10] suggested using deep learning (DL) techniques, seeking a broader and more detailed dataset. They also consider the possibility of modifying the questionnaire to make the responses in a suitable format, to increase the number of attributes used, and they suggest the inclusion of questionnaires from organizations such as the WHO (World Health Organization) related to stress and mental health. They also suggest formulating a homogeneous scale to assess stress levels. The article [11] mentions that people with common mental illnesses usually do not seek medical help, which makes attempts to monitor them to create opportunities for early intervention extremely difficult. The documented use of continuous digital monitoring to reach people with common mental illnesses among communities was noted as a strategy with some potential. At the same time, the limitations of monitoring systems based on assessments of mental health at specific points in time, on the basis of self-assessment and control by an expert, have been highlighted. These concern [11]:

- Impact of memory problems;
- Possibility to perform only in limited time windows;
- Possibility to perform only under controlled conditions;
- Frequent requirement for the patient to move to a medical setting in order to receive a diagnosis.

This raises further issues:

- Inability to assess the impact of interaction with the environment in the context of the mental state in real time;
- This undermines progress towards understanding and classifying mental illness and its treatment.

The authors also compare the use of mobile phones, in the context of dedicated solutions and solutions based on already available applications and devices. They point out that the second method has greater potential, as it significantly reduces costs and the risk of behavioral deformities associated with traditional forms of behavioral research [11]. In particular, they highlighted activity-tracking apps and wearable devices, which have received little attention in the context of research [11]. The study involved 53 of 120 recruited Australian volunteers aged between 18 and 25 years. They provided data in the form of a detailed health and lifestyle questionnaire and access to recorded information on activity-tracking apps. The Depression and Anxiety Scale-21 (DASS-21), which examines depression, anxiety and stress, was used to assess mental health. In addition, data on the duration of daily activities were included as a key point of interest. These were determined using data from miniature motion sensors, including location-based accelerometers, which were collected by various connected applications and fed into a cloud-based API, from where they were then stored in a database [11]. Based on the DASS-21, it was found that those monitored had symptoms of depression, anxiety and stress at intermediate levels. In contrast, the apps or devices that were linked to the API for the study were several:

- Fitbit;
- Garmin;
- Healthkit;

- Misfit;
- Moves;
- Myfitnesspal;
- Strava [11].

Based on the data collected, it was discovered that:

- Examined daily activity time received from wearable devices was greater than that derived from the mobile phone app;
- Of the 43 participants from whom at least three daily activity observations were obtained, 11 of them had at least 20% missing data between the first and last observation, but this did not show a relationship with DASS-21 scores;
- For the remaining 32 participants, entropy techniques were used, which initially showed no significant relationship between data and DASS-21 scale scores. It was not until splitting into two equal groups in relation to the amount of data that a significant, positive correlation was detected between the DASS-21 anxiety subscale and entropy in those with more data [11].

The authors [11] point to the lack of standardized systems for continuous mental health monitoring, which, together with continued monitoring in specific time windows, has contributed to the escalation of the problem. They note that people with mental health conditions are generally willing to share information from their mobile phones to help with research into these conditions, including serious illnesses. The authors present their work as a proof of concept for continuous mental health monitoring of mental health, but note the challenges of privacy, assessment and clinical integration and inclusion that would need to be addressed before it is more widely accepted. Another article [12], which deals with the determination of a voice-based mental health indicator using a mind-state observation system, explores the validity of such an approach. It draws attention to the huge cost of mental illness in developed countries and the need for early detection technology for depression and stress. Light is also shed on the current state of screening methods in the context of mental illness, including general health questionnaires (questionnaires including the General Health Questionnaire (GHQ) or the Beck Depression Index (BDI)). The effectiveness of such approaches in assessing disease conditions in the early stages was highlighted, and the problems of reporting bias, i.e., the effect of consciously or unconsciously under- or overestimating a patient's self-report, as well as the problem of reduced detection rates of mental illness in organizations with established hierarchies, were also noted. The authors of [12] report on their active research and work on voice-based mental health estimation. They list additional advantages of this approach:

- Ease of application;
- Possibility to monitor day by day, which conventional methods do not allow.

They have developed a software development toolkit (SDK) called MIMOSYS (https://medical-pst.com/en/products/mimosys/ accessed on 11 September 2023), whose features include:

- Recording a voice from a microphone;
- Analyzing this voice;
- Determining a health indicator based on this.

To enable daily monitoring, the authors developed a mobile app using MI-MOSYS. The aim of the study was to compare the indicator defined in the app with the BDI indicator. The study was carried out with the support of the local authority, which provided mobile phones with the mobile app installed for 50 company employees. The test participants had to record their voices by reading out ready-made phrases and talking using the device they were given. In addition, a BDI test was conducted at the beginning of the experiment. The voice analysis was based on the fact that people with mental illness show changes in the expression of emotions and changes in the proportions of the components of the voice. The four components hidden in the voice—anger, sadness, joy and calmness—were calculated from the characteristics of the recorded voice. In addition, the degree of excitement of the

respondent was determined. Taking these values into account, a short-term and a medium-term index of psychological well-being was determined, the latter based on short-term indices collected over a two-week period. As a result of the experiment, the correlation value was determined to be negative, with a value of 0.208 for the short-term value and 0.285 for the medium-term value. A lower correlation coefficient value was obtained for telephone calls, below 0.2 [12]. For the optimal cut-off, the following values of sensitivity, specificity and accuracy were obtained when analyzing the ROC curve:

- 0.795; 0.643; 0.660 for the short-term indicator;
- 1.000; 0.605; 0.646 for the medium-term indicator [12].

In the context of this research, the weak negative correlation between the indices from the app and the BDI was understandable, as a lower mental health index was associated with a higher rate of depression. Finally, the performance of the method in distinguishing between individuals with a high BDI was shown to confirm the appropriateness of the method. The efficiency of data accumulation was also noted, and furthermore, the results indicated that such a system could complement routine screening. However, the authors have set their sights on the commercialization of the product, as they do not disclose details in the form of the algorithms used or the scheme of operation of the system. Furthermore, it is not possible to download this toolkit without first contacting them via a form, which presumably means that it is made available for a fee. In addition, the library (Sensibility Technology) underpinning this software is also unavailable.

In [13], mental health before and during the COVID-19 pandemic was compared using a large probability sample from the UK population. The coronavirus and the methods used to slow its spread had a serious impact on people's livelihoods, incomes and debts, and was associated with serious concerns about an uncertain future. The authors of this publication [13] drew attention to the limited research on mental health during the pandemic, due to problems such as:

- Use of incomplete samples;
- Use of unverified or modified assessment tools;
- Lack of comparable pre-pandemic data to measure change.

Their study [13] was based on a large-scale survey conducted since 2009, including people aged 16 years and older. In addition, invitations to participate in the COVID-19 online survey were sent to participants in the last two series of surveys via emails, text messages and even letters. The pre-pandemic health assessment was based on data collected since 2014, and the data included results from the GHQ-12 questionnaire (a valid tool for assessing general mental health problems in the past two weeks, particularly effective in large-scale surveys). This scale was scored in two ways, the first based on a mean value and the second based on a binary threshold above which individuals were judged to have a significant level of mental health problems. The rating scale of this questionnaire for each question ranged from 0 to 3 (from no deviation to significant deviation). The authors [13] also carried out analyses by gender, age ranges, geographical location, or looking at the data from an ethnic perspective. Estimates of total annual income, employment status, living with a partner, age of the youngest child in the family were also analyzed, and a group of people at risk and those involved in COVID-19 was identified. Years with a small number of observations were excluded from the study, which may have led to less accurate estimates. Changes in mental health were also assessed using regression [13]. These models only included people for whom data from both the COVID-19 survey and at least one pre-pandemic data set were available, therefore 16- and 17-year-olds were excluded from this section. The value of the GHQ-12 index was constructed during the pandemic and placed in a time-variable model where average scores were used as the baseline, instead of using a binary index, as this would affect the statistical power of the results and their generalization. The final model included the following factors:

- Age;
- Sex;

- Family income;
- Employment status;
- Living with a partner;
- Presence of risk factors [13].

Various patterns related to variables have been detected, including [13]:

- Higher GHQ-12 scores in women;
- Higher scores in younger age groups;
- Slight differences in ethnicity (apart from the difference between Asians and white British—Asians scored higher);
- Slightly lower results were recorded outside cities;
- Higher scores in low-income families;
- Unemployed and professionally inactive people scored higher than employed and retired people;
- People without a partner and with young children had higher scores, as did the risk groups;
- Significant increase in average scores was noticed comparing the state before and during the pandemic [13].

The authors present their publication as one of the first in their country to measure the impact of the pandemic on the mental health of the population. The increase in mental health problems was not even among the designated groups. However, towards the end, they conclude that the increase was not significant, but point out the need for further studies spread over time, even postponed by half a year. They note that although GHQ-12 is a screening tool, it is not a clinical diagnosis. In the publication [14], it was mentioned that in the coming years, a radical change will be needed, consisting of attaching the patient's mental health profile to provide him with better treatment and help him recover faster. It was also noted that there has already been discussion about how medical predictive analytics could revolutionize healthcare globally. Factors affecting mental health include:

- Globalization;
- Pressures in the workplace;
- Competition [14].

The authors of [14] claim that the K-nearest neighbor's method, the naive Bayes classifier, or regression can be used to build the model. In their approach to identifying mental health, they used classification and clustering algorithms. They note the need for early diagnosis of deviations in mental health. The WHO report urged the nations of the world to harness the power of knowledge and technology to tackle mental health. They list some of the mental health assessment tools:

- Questionnaires;
- Sensors of wearable devices;
- Biological signals [14].

They also mention work on statistical relationships between mental health and other parameters, including:

- Educational achievements;
- Socioeconomic achievements;
- Satisfaction with life;
- Quality of interpersonal relations;

They also list various assessment methods [14] appearing in other works:

- Regression analysis;
- K-nearest neighbors method;
- Decision trees;
- Support vector method;
- Fuzzy logic;
- K-means method [14].

In their work [14], they started the analysis with clustering in order to better understand the data—obtaining certain groups, however, without any interpretation. They list and describe commonly used clustering methods:

- K-means;
- Hierarchical;
- Based on density;
- And their variants [14].

In addition, they presented frequently used indicators for validating clustering and applied the concept of the Mean Opinion Score (MOS) scale, used for subjective quality assessment. Their questionnaire consisted of 20 questions, posed to two populations: the first included 300 people aged 18 to 21, and the second 356 people aged 22 to 26. The rating scale for each question was five-point, from 1 (almost never) to 5 (almost always). The division into a set of training and test data was in the ratio of 80:20. In terms of validity, the best of all models were: bagging and random forest (0.90), slightly worse support vectors and K-nearest neighbors (0.89), and even worse logistic regression (0.84) and decision tree (0.81). The worst result was achieved by the naive Bayes classifier (0.73). It should be noted that the bagging algorithm uses multiple decision trees, trained on the basis of subsets of data selected by sampling with return. The remaining, undrawn data becomes the testing set. For already-built tree models, voting is used to get the final answer. The authors [14] pointed out that the quality of the features affects the reliability of the produced models, and they also propose the use of a feature subset selection strategy to shorten the learning time, or fuzzy logic when the number of classes is increased. In addition, they propose recursive neural networks as a possible option for larger data sets, also ensuring high accuracy. The authors of the publication [15], on the other hand, note the lack of a global definition for positive mental health, presenting various approaches to this issue. They mention the observation that definitions of good mental health are, and should be, to some extent context-dependent. The Public Health Agency of Canada, mentioned by the authors of [15], refers to positive mental health as the ability to feel, think and act in a way that strengthens the ability to enjoy life and cope with the problems encountered. Keyes describes it in a slightly different way, suggesting a definition of the syndrome of signs of positive feelings and positive functioning in life. The authors [15] note that a positive state of mental health is not synonymous with the absence of mental illness. This is the short version of the Mental Health Continuum (MHC), based on the concept of two related but distinguishable dimensions. The authors cite successful tests of this scale in countries such as Poland, Italy, Brazil and the United States. Many indicators of positive mental health have been identified in populations, including aspects such as general health, physical activity, sleep, substance use, violence or discrimination. For young people, factors such as relationships with peers or support from teachers are particularly important. Similarly, income, employment and place of residence were positively associated with good mental health. In their study, the authors [15] examined 5399 students from grades 8 and 10. All of them were willing to answer questions, and 92% of students answered all of them. The questionnaire used in the study was based on the Swedish version of the Survey of Adolescent Life in Vestmanland, which also included a short version of the MHC and other questions related to general health, substance abuse, exposure to technology, school life and socioeconomic background. Changed the wording of several questions to better fit the Chinese context. The data obtained were analyzed using SPSS 22 software, using multivariate logistic regression, likelihood ratios and 95% confidence intervals for the analysis of variables related to positive mental health as a dependent variable. In the beginning, the collinearity of the variables was checked by Spearman's correlation analysis. Further, insignificant indicators were dropped until the model was statistically significant. Nagelkerke's Pseudo-R2 statistic and model fit were also calculated. Their research [15] extends knowledge about the prevalence of positive mental health among Chinese minors, as well as about the indicators of positive mental health. As a result, information was obtained that the surveyed group of Chinese people was significantly healthier in terms

of mental health than in similar studies in other countries. The authors acknowledge that their study covered only one city in China, so further research in different regions will be needed. On the other hand, the authors of the publication [16] on economic difficulties and reported mental health problems during the COVID-19 epidemic point to the problem of isolation increasing the risk of loneliness, or the need to assess the links between the labor market and mental health, also in order to understand the impact of the pandemic on existing the socioeconomic inequalities. Their considerations [16] include factors related to changes in workload, income decline and job loss, as well as three mental health issues:

- Depression;
- Loneliness;
- Fear for your health [16].

The data came from employee surveys in Italy, Spain, the Czech Republic, Slovakia, the Netherlands and Germany from March and April 2020. The research also took into account the International Socio-Economic Index (ISEI). It expresses the relative position of the profession in the labor market, on a scale of 10 to 89 points. During the analyses [16], it was noted that occupations with an ISEI index below 30 points were characterized by a much higher risk of economic difficulties—about twice as high as medium and high-rated occupations (ISEI up to about 80 points). In addition, freelance and self-employment increased the likelihood of a reduction in workload by more than 32 percentage points, a decrease in income by 42 percentage points, and a loss of a job by just under 20 percentage points, compared to typical workers. Similarly, in the comparison between employees and employers, reductions in workload and income were more pronounced in the first group. In the final part of the work [16], they point out that the indicators used by them are not clinically confirmed, which makes it impossible to compare them on an equal basis, but they are an assessment of feelings about mental health. In addition, they consist of single questions, which makes them a non-detailed assessment of mental health. The authors explain that this is due to the data in the questionnaires not being designed to capture mental health, so researchers have had to rely on crude indicators. On the other hand, in the paper [17] attention was drawn to incomplete or partial evidence of the connection between mental illnesses and work. Therefore, the authors assumed that the mental health of an individual depends on characteristics such as:

- Personality;
- Sex;
- Own results at work;
- Loss of a job by a family member [17].

They developed [17] two models, one for the issue of the impact of job loss by a partner on the spouse, and the other describing the effects of parental job loss on underage children. They also sought to limit biasing effects in their study, based on data from around 7700 Australian households. The data consisted of responses to the Household, Income and Labor Dynamics in Australia (HILDA) survey. In order to develop two models, two separate data samples were created [17]—one for married couples, the other for parent–child pairs. Part of the data included answers to the Self-Completion Questionnaire (SCQ), which the researchers used in both the first data sample and the second. The MHI-5 (MHI—Mental Health Inventory) was used as the output variable [17], consisting of five questions on a 6-point scale. These questions were as follows:

- Were you a nervous person?
- Have you felt so down that nothing could cheer you up?
- Did you feel calm and composed?
- Have you felt depressed?
- Were you a happy person? [17].

The scores on this scale ranged from 0 to 100, where the lower the value, the worse the mental health. As a result of these studies [17], it turned out that the impact of losing a wife's job had no greater effect on husbands, while wives whose spouses lost their jobs had

between 2 and 2.7 lower scores than women whose husbands still had jobs. However, the authors, taking into account other factors, indicate that this is not a statistically significant result. It was only when differentiating between groups with persistent unemployment, financial stress and dissatisfaction with relationships that a significant effect of losing a job by husbands was found. They found that continued unemployment caused a significant decline in mental health between studies and that the financial stress situation did not significantly contribute to worse mental health, while both women and men experienced worse mental health as dissatisfaction with their partner increased compared to previous answers. However, looking at the results [17] regarding the mental health of children after the loss of a job by one of the parents, it did not have a significant impact on its deterioration. A drop of 6.6 points was recorded when the mother was unemployed between examinations, which has a much higher impact than was observed for other variables. Comparing the mental state of boys and girls, it was shown that the deterioration of mental health was greater in girls, especially when the mother was unemployed. However, in the work [18], the impact of natural disasters on the mental health of minors is compared with their peers who have not experienced such events. Their study uses data on students from two Canadian cities located in the same province (Fort McMurray and Red Deer). In the surveys conducted in these cities, six questionnaires common to both studies were used, including:

- Patient Health Questionnaire, Adolescent version (PHQ-A);
- Hospital Anxiety and Depression Scale (HADS);
- CRAFFT questionnaire;
- Tobacco Use Questionnaire;
- Rosenberg's self-esteem scale;
- Kidscreen questionnaire [18].

The authors [18] performed a statistical analysis based on these questionnaires, and also compared the percentage odds of:

- Depression;
- Thoughts of suicide;
- Medicines;
- Using alcohol/stimulants;
- Tobacco use;
- Any of the options: about depression, about fears or use of alcohol/stimulants.

An additional limitation was the use of only complete answers for each measure, i.e., without omitted questions. A comparison [18] of indicators between the two regions found significant differences in 8 out of 12 measures of mental health status. The rates of possible depression were significantly higher in the city that experienced a natural disaster, as were those for suicidal thoughts and tobacco use. On the other hand, the self-esteem and quality of life scales (Rosenberg and Kidscreen, respectively) were much lower, but this is related to the nature of their questions. The conclusions [18] include the observation that this research reinforces the need for policies and programs to care for mental health among minors, especially after natural disasters, in order to reduce their vulnerability and build a positive state of mental health. They also note that it would be useful to compare these studies with data for post-traumatic stress symptoms from both cities, as the authors did not have such data from the city of Red Deer. They also indicate that minors are very vulnerable to the adverse impact of natural disasters. Summing up the studied literature, it can be noted that these are extremely diverse studies, they address many aspects related to mental health indicators, both from the side of positive and negative mental health. In addition, a variety of approaches were used, including voice data analysis, conducting surveys using many different questionnaires, random forests, bagging algorithm, support vector method, K-nearest neighbor's method, and statistical analysis. However, it should be borne in mind the need to expand research in the search for more effective algorithms that can be used in this area.

The proposed solution can be used in a prototype preventive mental health medicine system (Figure 2) for healthy people to monitor and detect the first symptoms of chronic stress and burnout as early as possiblebased on a combination of a generic standard and a dynamic standard generated directly from the data set. Given the second opinion offered by the ML system, it will support the activities of primary care physicians and psychology and psychiatry specialists in their daily efforts to provide early diagnosis and treatment of this group of conditions and will allow the selection and application of prevention and, if necessary, minimize the duration of potential therapy and reduce its cost [19].
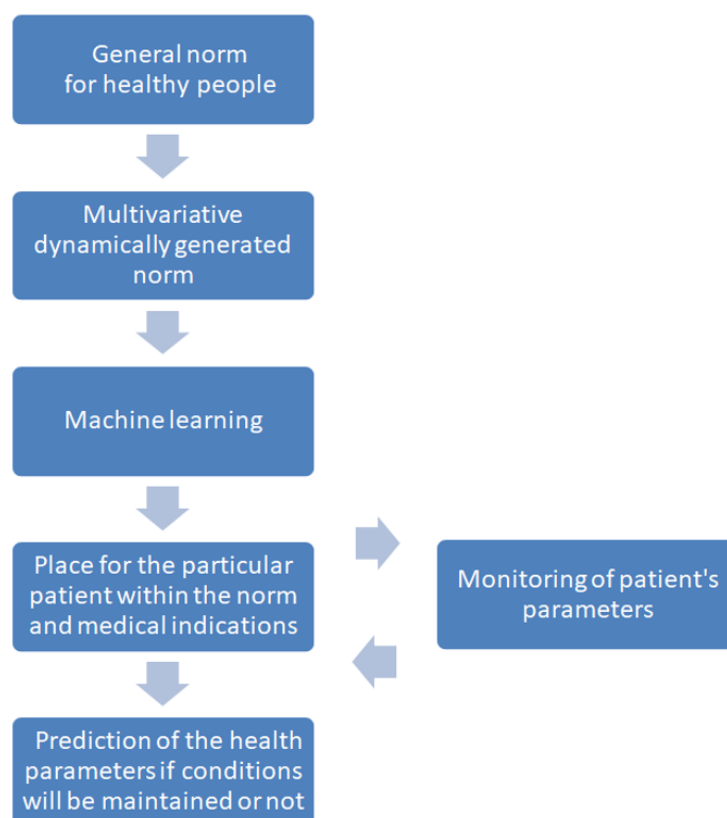
Figure 2. Prototype of preventive mental health medicine system [19].

Novelty and contribution lie in the application and matching of ML methods to the form and characteristics of test data describing chronic stress and job burnout. Pre-selection of methods and their initial facilitated matching to presumed criteria is key, which will support the development of preventive mental health medicine systems.

The research aims to determine a virtual indicator of mental health using selected ML algorithms, as well as to determine their effectiveness in this task by checking the learning time, operation and accuracy. In addition, research hypotheses will be verified, i.e.,:

- Choice of the ML method affects the regression accuracy, learning time and running time;
- Differences in accuracy are relatively small—up to about 10 percentage points difference between methods.

2. Materials and Methods

2.1. Material

The results of 99 patients (36 women and 63 men, mean age 27.93, SD = 4.64, mean seniority 3.78, SD = 2.94) with suspected chronic stress and burnout were analyzed using ML (Table 1).

Table 1. Data set distribution.

Parameter	Mean	SD	Min	Q1	Median	Q3	Max
PSS item 1	2.96	0.79	1	2	3	4	4
PSS item 2	3.14	0.74	2	3	3	4	4
PSS item 3	2.87	0.92	1	2	3	4	4
PSS item 4	2.66	1.05	0	2	3	3	4
PSS item 5	3.06	0.65	1	3	3	3	4
PSS item 6	2.90	0.85	1	2	3	3	4
PSS item 7	3.08	0.97	1	3	3	4	4
PSS item 8	2.67	0.90	0	2	3	3	4
PSS item 9	2.94	0.71	1	3	3	3	4
PSS item 10	2.49	0.93	1	2	2	3	4
MBI item 1	3.27	1.96	0	2	3	5	6
MBI item 2	2.73	1.73	0	2	3	4	6
MBI item 3	2.49	1.70	0	1	3	3	5
MBI item 4	2.24	2.32	0	0	1	5	6
MBI item 5	1.50	1.68	0	0	1	3	6
MBI item 6	1.53	1.48	0	0	1	3	6
MBI item 7	3.37	1.78	0	2	3	5	6
MBI item 8	1.69	1.68	0	0	1	3	6
MBI item 9	2.86	2.57	0	0	3	6	6
MBI item 10	1.56	1.35	0	1	1	3	6
MBI item 11	2.09	1.55	0	0	3	3	6
MBI item 12	2.55	1.66	0	1	3	3	6
MBI item 13	2.09	1.52	0	1	2	3	6
MBI item 14	2.17	1.86	0	1	1	3	6
MBI item 15	2.36	2.03	0	0	2	4	6
MBI item 16	1.68	1.77	0	0	1	2.5	6
MBI item 17	2.76	2.04	0	1	3	3	6
MBI item 18	1.64	1.45	0	0	1	3	5
MBI item 19	2.56	1.95	0	1	3	3	6
MBI item 20	1.87	2.23	0	0	1	3	6
MBI item 21	1.21	1.48	0	0	0	3	4
MBI item 22	2.43	1.45	0	2	3	4	6
SWLS item 1	4.08	0.93	2	4	4	5	5
SWLS item 2	3.24	1.53	1	2	3	4	6
SWLS item 3	3.30	1.66	1	1	4	5	6
SWLS item 4	3.20	1.66	1	2	2	4	6
SWLS item 5	2.51	1.59	1	1	2	4	5

Mental well-being data were used, including people's gender, age, length of service and their responses to the three questionnaires: Perceived Stress Scale (PSS), Maslach Burnout Inventory (MBI) and Satisfaction with Life Scale (SWLS).

The subject of the study was data from a set of 99 people, information about which was divided into 4 subgroups, each in a separate MS Excel sheet: "Patient data", "PSS10", "MBI", and "SWLS". The first of the above sheets includes the patient's gender, age and work experience. The second sheet contains answers to 10 questions from the PSS set, on a scale of 0 to 4, where 0 corresponds to "never", 1—"almost never", 2—"sometimes", 3—"quite often", and 4—"very often". The third sheet contains answers to 22 questions from the MBI set, on a scale of 0 to 6, where 0 corresponds to "never", 1—"several times a year", 2—"once a month", 3—"severaltimes a month", 4—"once a week", 5—"several times a week", and 6—"every day". The fourth sheet contains answers to 5 questions from the SWLS set, on a scale of 1 to 7, where 1 corresponds to "strongly disagree", 2—"disagree", 3—"slightly disagree", 4—"neither agree nor disagree", 5—"agree slightly", 6—"agree", and 7—"strongly agree". Based on these four sheets, a CSV (Comma Separated Values) file was created, which was used in the application due to the inability to directly load an Excel file with the .xls extension, also taking into account the available NuGet packages—they are satisfactorily documented for use in the project. This CSV file uses a semicolon (;) as the

delimiter, which has been included in the app as the default delimiter value. The total is based on all answers from PSS, MBI and SWLS sets. All but the first column of the CSV file contain numeric values, while the first column can only contain two options: M (Male) or F (Female).

The study was approved by the Bioethics Committee No. KB 391/2018 at the Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń. Each participant in the study gave informed consent.

2.2. Methods

Two languages were used to develop the application: C# in .NET and Extensible Application Markup Language (XAML), whereby:

- C# language was used to describe the actions performed by the program;
- XAML was used to develop the layout of the user interface in a Universal Windows Platform (UWP) application, along with the naming of elements (which allows them to be used in C# as variables), or the binding of events to specific functions in the code behind the interface (code behind).

A number of ML models based on Stochastic Dual Coordinate Ascent (SDCA), limited-memory Broyden–Fletcher–Goldfarb–Shanno, Online Gradient Descent, etc., were built based on a clinical dataset (PSS, MBI and SWLS) and compared based on criteria in the form of learning time, running time during use and regression accuracy. The rationale for choosing these particular algorithms lies in their popularity and the authors' previous experience and previous research on measuring long-term stress and burnout using the aforementioned group of tests and AI [19–22]. Knowledge in the area of matching AI/ML tools for the analysis, inference and prediction of stress and burnout measurements is still nascent and no computational or theoretical basis can be cited as yet.

The predicted value was a virtual mental health index.

The data set has been divided into a training set (70% of samples) and a test set (30% of samples).

SDCA algorithm is a linear algorithm, meaning that it generates a model that calculates results based on a linear combination of the input data and a set of weights. The model weights are those parameters that are determined during training. In the general case, linear algorithms are scalable, fast and have a low cost during training and during prediction. This class of algorithms goes through the training dataset many times [23]. It is devoid of parameters for manual tuning and has a clearly defined stopping criterion. This algorithm has good implicit performance. It combines some of the best features, such as:

- Possibility of streaming learning, i.e., operating on data without having to put it all in memory at once;
- Achieving satisfactory results with a small number of circuits through the entire data set;
- Not wasting computing power on zeros in sparse datasets [24].

It should be borne in mind that the results obtained with this algorithm are dependent on the order of the training data, but the solutions obtained can be treated as equally good between different executions of the algorithm [25]. This algorithm is a stochastic version of DCA. The basic version of the algorithm (DCA) performs optimization on a single variable in each iteration without affecting the others. The SDCA version of the algorithm performs a pseudo-random selection of a double coordinate for optimization based on a uniform probability distribution [26].

LBFGS is an abbreviation for limited-memory Broyden–Fletcher–Goldfarb–Shanno, an optimization algorithm based on BFGS, but using limited Random Access Memory (RAM) [27,28], as it does not store a matrix approximating the inverse of the Hessian $\nabla 2$ f(x), instead using an intermediate approximation [28,29]. The calculation is based on an initial approximation and an update rule that models local curvature information [27,28]. The original Broyden–Fletcher–Goldfarb–Shanno method, called full BFGS (BFGS), pro-

posed by these four authors in 1970, keeps the aforementioned matrix in memory, whose computational cost of updating is high, of the order of O(n2) [28,29]. As for the convergence of the BFGS method, if the function has a continuous second derivative and the function is strongly convex, the sequence of successive values of x_k tends towards the global minimizer, and furthermore, when it is assumed that the Hessian satisfies the Lipschitz condition, the rate of convergence is super linear [28,29]—i.e., faster than linear. The convergence of the LBFGS algorithm depends on the quality of the Hessian approximation, which is difficult to achieve, and it has been observed in numerical observations that an appropriate guess of the initial Hessian has a significant impact on the search direction and convergence [27,28].

The Online Gradient Descent (OGD) algorithm is a variation of the Stochastic Gradient Descent (SGD) method used for online training—i.e., training by learning concepts incrementally by processing examples from the training set one at a time, one after the other, with the algorithm not storing the last occurrence after each update, but based on the next sample [29,30]. SGD uses an iterative technique based on error gradients, in addition to providing the ability to update the weight vector using the average of the observed data vectors as the algorithm progresses [31]. SGD is popular for its simplicity, computational efficiency, or convergence independent of the training dataset, and the performance of DL methods depends heavily on this algorithm. However, it is susceptible to the effects of noisy data, especially noticeable in robotics, where robots do not have the capacity to collect enough data to negate these effects [32].

Three questionnaires were used to determine a virtual mental health index: PSS, MBI and SWLS. Data from these questionnaires were used in the application to train the models and determine metrics and statistics.

PSS is a scale developed by Cohen, Kamarck and Mermelstein in 1983, which aimed at respondents' self-assessment of the unpredictability of their life, their lack of control over it and the overload they feel. The original version has fourteen general questions on a four-point scale, and the final score is obtained by reversing the scale for positively valenced questions and then adding up the scores for all questions. In addition, two shorter versions of the scale have been developed, the ten-question scale used in this work, as well as a four-question scale [33]. Research on this instrument is massively carried out all over the world, including China, Ethiopia, Iran and Greece, and the results indicate that this scale can be relied upon to be used in these countries. To validate the scale, Cohen studied the responses of people of different ages, both genders and a variety of racial backgrounds [34]. Similar information is presented by the authors of a Czech study, where they briefly describe that all versions of the scale had previously been compared in a variety of cultural and linguistic contexts and that these researchers agreed that the ten-question scale was at least comparable to or better than the original version in terms of internal consistency while noting a significant decrease in reliability of the four-question version, which was attributed to it simply being too short [33].

The MBI (Maslach Burnout Inventory) was developed by Christina Maslach and her team. In her article, she explains the concept of professional burnout (burnout)—a syndrome of emotional exhaustion and cynicism often found in people who work with people, with a key component being the increased sense of emotional exhaustion mentioned earlier. It indicates that with the depletion of their emotional resources, employees begin to feel that they are not able to give their best; furthermore, they develop negative, even cynical attitudes and feelings about their clients. The two aspects seem to be linked, and a tendency to evaluate oneself negatively, especially in relation to one's work, not feeling satisfied with one's achievements, is mentioned as a third effect related to professional burnout [34]. Occupational burnout is characterized by high levels of emotional exhaustion, dehumanization and low feelings of personal fulfillment. In addition, they point out that occupational burnout and depressive states are related, but they are not the same concepts, i.e., their characteristics do not overlap and thus cannot be used interchangeably [35]. The version used in this thesis consists of three groups of questions regarding these issues:

emotional exhaustion (nine questions), sense of personal accomplishment (eight questions) and dehumanization (five questions).

In terms of SWLS acceptability, reliability, validity as well as gender independence have been demonstrated, as indicated by the authors of the article [36]. The scale was first presented in 1985 and was summarized as narrowly focused on the issue of overall satisfaction with life, without addressing issues such as loneliness or positive affect [37]— which is described as the feeling experienced when a certain goal is achieved, or a source of danger is averted, or the person is satisfied with the current state of affairs [30]. It was developed as a response to a number of scales that contained only one question, and to scales that went beyond life satisfaction. The process of shaping the questions in this set began with a list of 48 questions, and, after eliminating questions about affect and questions with a Factor Loading of less than 0.60 and omitting those with a high degree of similarity, yielded five questions [36], scored on a scale of 1 to 7, which in effect generated a score range of 5 to 35.

Test results and calculations were recorded in an MS Excel spreadsheet.

Statistical analysis was performed using Statistica 13 (StatSoft, Tulsa, OK, USA). The Shapiro–Wilk test was used to check the normality of the distribution of the studied data. The *p*-value was set at 0.05. Where possible, analyzed values with distributions close to normal were presented as mean values and standard deviation (SD). The analyzed values with distributions different from the normal distribution were presented using the minimum value, the lower quartile (Q1), the median, the upper quartile (Q3) and the maximum value.

Selected ML algorithms were compared on the basis of:

- Metrics: mean absolute error, mean squared error, mean squared error, coefficient of determination;
- Learning time, expressed in milliseconds: minimum, average, maximum;
- Prediction time, expressed in milliseconds: minimum, average, maximum.

For each of the compared values, the best algorithm was determined, choosing the one with:

- Minimum value for learning times, prediction, mean absolute error, mean squared error;
- Maximum value for the coefficient of determination.

In a similar way, the worst algorithm in terms of a given criterion was determined, this time by:

- Maximum value for learning times, prediction, mean absolute error, mean squared error, mean squared error,
- Minimum value for the coefficient of determination.

In order to make more accurate use of the software's capabilities, each learning was performed four times in order to make the resulting parameters more meaningful and, in addition, each time the steps were carried out on a new occurrence of the application. To compare the algorithms, we selected the best hyperparameters for each optimizer and for each data set using a validation procedure with a learning set and a validation set. On each data set, for each hyperparameter, we calculated the accuracy after a certain number of epochs for a range of values and a certain validation set. The study considered the following hyperparameters of each of the algorithms tested:

- SDCA: c (regularization strength) and stopping time;
- LBFGS: solver, penalty, max_iter, c, tol, fit_intercept, intercept_scaling, class_weight, random_state, multi_class, verbose, warm_start, and l1_ratio;
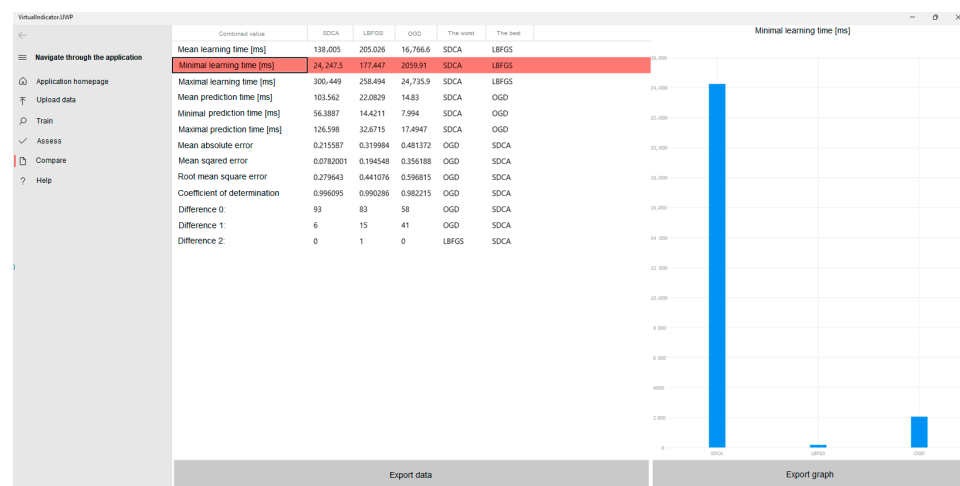- OGD: learning rate and diameter of the decision set.

The same data set was also analyzed automatically using ML.NET (Visual Studio 2022, Microsoft, Redmond, WA, USA).

3. Results

The algorithm with the highest accuracy was Stochastic Dual Coordinate Ascent, but although its performance was high, it had significantly longer training and prediction times (Figure 3a).

(**a**)

(**b**)

(**c**)

Figure 3. (**a**) General comparison of metrics, training times and predictions, (**b**) comparison of selected metrics and (**c**) assesment of models (three columns on right side).

The fastest algorithm looking at learning and prediction time, but slightly less accurate, was the limited-memory Broyden–Fletcher–Goldfarb–Shanno (Figure 3b).

The first criterion considered was the model learning time, expressed in milliseconds. The average, minimum and maximum values were taken into account. Both the average, minimum and maximum times were the longest for the SDCA model and the shortest for the LBFGS model. This means that the SDCA model performed the worst in this ranking, and the LBFGS model performed best. It should be noted that while for the LBFGS and OGD models, the difference between their maximum and minimum values was relatively small (about 4% of the average, both for LBFGS and OGD), for the SDCA model it was about 64% of the average. Another criterion was the prediction time for the entire data set, expressed in milliseconds. The average, minimum and maximum values were taken into account. This time was the lowest for the OGD model, but it differed only slightly from the LBFGS model, both models reached a time slightly above 1 ms. On the other hand, the average time for the SDCA model was about 44 times longer than for the OGD model, and again there were larger differences between the maximum and minimum values for the SDCA model (approximately 18% of the mean value). The average absolute error was the lowest for the SDCA model and amounted to about 0.216, while it was the highest for the OGD model, amounting to about 0.481 (more than twice as much). On the other hand, for the LBGFS model, it was around 0.320, which corresponds to an increase of 48%. For this criterion, as well as for the mean squared error and the mean squared error, the best results were achieved by the SDCA model, and the worst by the OGD model. The ranking for the coefficient of determination looks similar, the value of which is closer to 1, the better for the model. The last lines of the comparison show the number of occurrences for which the absolute value of the difference between the rounded prediction and the value from the dataset was 0, 1 or 2, respectively. Looking at the difference equal to 0, the best result was obtained by the SDCA model, and the worst by the OGD model. For a difference of 1, the best result was obtained by the SDCA model (6 occurrences), and the worst by the OGD model (41 occurrences). However, for the difference equal to 2, there was one such occurrence for the LBFGS model (Figure 3c).

In this particular problem of determining a virtual mental health index, all three models considered achieved comparable final results. Based on the criterion of model learning time, and considering other factors (e.g., prediction time), the LBFGS model would be the best choice. On the other hand, looking at metrics in the form of, among other things, mean absolute error or coefficient of determination, the SDCA model, whose biggest drawbacks are learning time and prediction time, would prove to be the best choice. Although the OGD model achieved the best prediction time, it achieved the worst of the results when looking at the metrics.

Looking at the results obtained, the differences between the ML methods used are clearly visible, especially for learning time and metrics. Furthermore, bearing in mind that the accuracy of the model increases as the value of the coefficient of determination approaches one, the differences between the methods amounted to a maximum of around 1.4 percentage points, looking at the difference between the maximum and minimum value in relation to the maximum possible value, i.e., 1 (which can be understood as 100%).

We have compared the aforementioned results with automated analysis using ML.NET results (249 models checked, Tables 2 and 3).

Table 2. Results of ML-based classification.

Parameter	Micro Accuracy (%)	Macro Accuracy (%)	Best Trainer
Gender	75.16	69.32	FastTreeOva
Age	71.24	62.82	FastTreeOva
Seniority	78.73	72.23	FastForestOva
Total pts.	17.29	14.79	LightGbmMulti

Table 3. Results of ML-based prediction.

Parameter	Accuracy (%)	Best Trainer
Gender		Not possible
Age	93.32	LbfgsPoissonRegressionRegression
Seniority	97.57	FastTreeRegression
Total pts.	97.42	LbfgsPoissonRegressionRegression

Despite the fact that the data lends itself to both prediction and classification, it has not been possible to find one algorithm that is good at everything—a thoughtful combination of different algorithms must be used in automated analysis.

4. Discussion

A comparison of the three ML algorithms showed small differences in regression accuracy (about 1.4 percentage points, or, according to the thesis, less than 10 percentage points), which, in relation to the work [10], which nevertheless dealt with the classification problem, but revealed differences in accuracy between six different methods of about 5.5 percentage points, probably means a small impact of the method used on regression accuracy or classification accuracy.

The results of the paper [14] are similar, where all the algorithms used, except for the naive Bayes classifier—which is the simplest one used and probably for this problem did not have a strong connection to reality—obtained accuracy differences of at most 9 percentage points. Looking at the proposal in that paper, continuing research would need to use a feature subset selection strategy so that the solution is based on the highest quality features. Applying such an approach successfully would mean a reduction in learning time and potentially an increase in model reliability. In addition, the inclusion of the patient's mental profile mentioned can be considered to have been done, as the data contains answers to a set of questions to assess the patient's mental health status.

When comparing with studies [13,16–18], it is important to note the lack of analysis of the impact of individual factors on the virtual mental health index, considering particular attributes such as age, gender, or length of service. This implies an opportunity for further research to be able to establish some trends, for example among different age groups, as in the article [6]. In addition, further data would have to be collected, not only more numerous but possibly also including the ISEI index, which expresses the relative position of the occupation in the labor market, as in the study [16]. Regarding the study [17], the dataset could be extended to include information on the dynamics of employment, or also the household of the person surveyed. Looking at the study [18], it would be valuable to assess the risk of problems such as depression, anxiety or the use of stimulants, which could be baseline variables for the trained models.

Referring to the work [11], which addresses the problems of assessing health status in discrete moments in time, mainly in terms of not being able to assess the impact of the environment on the patient in real time, one could use data from apps and activity monitoring devices of potential volunteers to derive models based on measured data. On the one hand, this would make it possible to assess mental health on a continuous basis, and on the other hand, it would make it independent of the patient's self-assessment.

As described in the paper [12], voice-based mental health determination, while it appears to be a promising solution, the authors did not present the ML methods used, which, combined with the commercialization of the developed library and system, does not allow for the extension of these analyses. On the other hand, the idea is intriguing, but in order to be realized, it would require an appropriate selection of libraries and ML methods, as well as the disposal of voice data, together with the determination of the patient's mental health status for these data.

On the other hand, the observation in the article [15] that a positive mental health status does not imply the absence of mental illness, which was taken into account in the Mental Health Continuum scale, whose tests in various countries have been successful.

This is something to bear in mind, as it happens that mental illnesses are able to be hidden, both consciously and unconsciously. It is also important to consider factors that are often indicative of a patient's mental state, such as their physical activity, sleep, use of stimulants, and relationships with peers in the case of adolescents or relationships with co-workers among adults.

In the study, learning time or prediction time is an evaluation criterion. The performers' tasks do not require real time, but with large databases and a large number of simultaneous system users, the value of this parameter can be very important.

It is noteworthy that a variety of tools have been used in these papers, whether in the form of questionnaires, such as the Depression Anxiety Stress Scale-21, Beck Depression Inventory or algorithms (logistic regression, K-nearest-neighbor method, decision trees, bagging, support vector method) and technology, including the Python language, Scikit-learn library, physical activity tracking mobile apps and wearable devices. In addition, many of these papers did not present the programming language used, making it impossible to make a technology choice based on them.

Key findings in the area of ML-supported human mental health analysis have shown that, despite the variety of tools that have been used in these papers, one leading approach is lacking, both in the selection of tests and in the selection of ML-based aggregation and analysis methods. This makes it difficult both to compare different approaches and to extract the best ones (based on common criteria) for further development and use in both simple predictive systems within preventive medicine and complex diagnostic and monitoring systems within more complex specialized studies. This results in the unique contribution of the current study compared to the existing literature, which includes how to aggregate test results into a virtual mental health index and how to select optimal ML methods for its further use providing a basis for further research, including for other groups of clinicians and researchers. Our experience to date shows that this element of technological support is lacking in clinical practice, hence interdisciplinary teams are needed for further research.

4.1. Limitations of Studies

Research on determining a virtual index of mental health using ML algorithms may encounter a number of limitations and challenges that should be considered:

- Lack of unequivocal measures of mental health (patients and healthy people)—mental health is a subjective concept and difficult to define unambiguously, which complicates the process of creating ML models;
- Population diversity—individual healthy individuals differ from each other in terms of mental health as well as in different life contexts, which makes general modeling difficult and it will be necessary to adapt models to different population groups;
- Lack of qualitative data—most of the available data is quantitative, which can hinder a fuller understanding of mental health;
- Lack of historical data—it is often important to consider the historical context of the patient's illness;
- Data privacy—mental health data are very sensitive, so it is necessary to maintain appropriate standards of data privacy and security;
- Cultural differences—Mental health can be understood and experienced differently in different cultures.
- Interpretability of models—understanding why a model made certain decisions can be a problem for mental health diagnosis and treatment;
- Importance of experts—ML models will not replace human expertise, but will only support it [38–42].

4.2. Directions for Further Research

Research on the determination of a virtual index of mental health using ML algorithms is an area that can bring many benefits in the field of health care and mental well-being,

as well as their objective, partially automated assessment and monitoring of changes. A summary of research directions that can be explored in this context is presented in Table 4.

Table 4. Directions of research on the virtual index of mental health with the use of ML algorithms [43–45].

Area	Description and Detailed Tasks
Data collection and analysis	The use of many different data sources, including multi-modal ones, such as behavioral data (e.g., online activity, phone calls), biometric data (e.g., heart rate, sleep monitoring), survey data, photos and videos, as well as test results collected automatically, etc.
Collaboration with field experts	Collaboration with physicians and mental health professionals can help understand the mechanisms and create and evaluate the effectiveness of models.
Ethics and privacy	The manner in which data are collected, stored, used and destroyed should comply with relevant regulations and ethical standards.
Data preparation	Data preparation may include data normalization, removal of erroneous, uncertain, incomplete and outlier data, coding of categorical variables, etc.
Selection of ML algorithms and hyperparameters	Selection and adaptation of algorithms and hyperparameters of models to a specific problem from among possible solutions, such as decision trees, neural networks, support vector machines (SVM) or clustering algorithms.
Evaluation/cross-validation of models	Define model performance metrics (accuracy, sensitivity, specificity, F1-score, Receiver Operating Characteristic (ROC) curves, etc.) and analyze model performance using them.
Interpretability of models	Understanding how the model makes its predictions (why the model made certain decisions).
Checking the learning time	Model training time can be a critical factor in clinical practice—it needs to be investigated how long it takes to train different models and whether this can be optimized.
	The model should be adapted to real-time operation (including learning on new patients) in order to be used in clinical practice.
Validation on a large sample of patients	The effectiveness of the models should be tested on a large sample of patients to ensure that the model generalizes well to different cases.

This research can be a long and complicated process, but it can have significant benefits in diagnosing, monitoring and managing patients' mental health [46,47].

5. Conclusions

The ability of ML to identify burnout using passively collected electronic health record (EHR) data and predict future health status with an accuracy of more than 70% (for some traits: more than 90%) accounts for the usefulness of this group of methods in daily clinical practice, which is worth developing.

The algorithms did not differ significantly from each other in terms of accuracy (about 1.4 percentage points) but differed more strongly in other parameters. The algorithm with the highest accuracy was Stochastic Dual Coordinate Ascent, but although its performance was high, it had a significantly longer training and prediction time. In contrast, the fastest algorithm looking at learning and prediction time, but slightly less accurate, was the limited-memory Broyden–Fletcher–Goldfarb–Shanno.

Findings from the study can be used to build larger systems that automate early mental health diagnosis and help differentiate the use of individual algorithms depending on the purpose of the system.

References

1. Asatryan, B.; Bleijendaal, H.; Wilde, A.A.M. Toward advanced diagnosis and management of inherited arrhythmia syndromes: Harnessing the capabilities of artificial intelligence and machine learning. *Heart Rhythm.* **2023**, *20*, 1399–1407. [CrossRef]
2. Kannampallil, T.; Dai, R.; Lv, N.; Xiao, L.; Lu, C.; Ajilore, O.A.; Snowden, M.B.; Venditti, E.M.; Williams, L.M.; Kringle, E.A.; et al. Cross-trial prediction of depression remission using problem-solving therapy: A machine learning approach. *J. Affect. Disord.* **2022**, *308*, 89–97. [CrossRef] [PubMed]
3. Hong, N.; Liu, C.; Gao, J.; Han, L.; Chang, F.; Gong, M.; Su, L. State of the Art of Machine Learning-Enabled Clinical Decision Support in Intensive Care Units: Literature Review. *JMIR Med. Inform.* **2022**, *10*, e28781. [CrossRef]
4. Lopez-Jimenez, F.; Attia, Z.; Arruda-Olson, A.M.; Carter, R.; Chareonthaitawee, P.; Jouni, H.; Kapa, S.; Lerman, A.; Luong, C.; Medina-Inojosa, J.R.; et al. Artificial Intelligence in Cardiology: Present and Future. *Mayo Clin. Proc.* **2020**, *95*, 1015–1039. [CrossRef]
5. Reid, J.E.; Eaton, E. Artificial intelligence for pediatric ophthalmology. *Curr. Opin. Ophthalmol.* **2019**, *30*, 337–346. [CrossRef] [PubMed]
6. Mentis, A.A.; Lee, D.; Roussos, P. Applications of artificial intelligence–machine learning for detection ofstress: A critical overview. *Mol. Psychiatry* **2023**, 1–13. [CrossRef]
7. Galatzer-Levy, I.R.; Onnela, J.P. Machine Learning and the Digital Measurement of Psychological Health. *Annu. Rev. Clin. Psychol.* **2023**, *19*, 133–154. [CrossRef] [PubMed]
8. Sutrisno, S.; Khairina, N.; Syah, R.B.Y.; Eftekhari-Zadeh, E.; Amiri, S. Improved Artificial Neural Network with High Precision for Predicting Burnout among Managers and Employees of Start-Upsduring COVID-19 Pandemic. *Electronics* **2023**, *12*, 1109. [CrossRef]
9. Adapa, K.; Pillai, M.; Foster, M.; Charguia, N.; Mazur, L. Using Explainable Supervised Machine Learning to Predict Burnout in Healthcare Professionals. *Stud. Health Technol. Inform.* **2022**, *294*, 58–62. [CrossRef]
10. Srinivasulu Reddy, U.; Thota, A.; Dharun, A. Machine Learning Techniques for Stress Prediction in Working Employees. In Proceedings of the 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 13–15 December 2018; pp. 1–4.
11. Knight, A.; Bidargaddi, N. Commonly available activity tracker apps and wearables as a mental health outcome indicator: A prospective observational cohort study among young adults with psychological distress. *J. Affect. Disord.* **2018**, *236*, 31–36. [CrossRef]
12. Hagiwara, N. Validity of Mind Monitoring System as a Mental Health Indicator using Voice. *Adv. Sci. Technol. Eng. Syst. J.* **2017**, *2*, 338–344. [CrossRef]
13. Pierce, M. Mental health before and during the COVID-19 pandemic: A longitudinal probability sample survey of the UK population. *Lancet Psychiatry* **2020**, *7*, 883–892. [CrossRef] [PubMed]
14. Srividya, M.; Mohanavalli, S.; Bhalaji, N. Behavioral modeling for mental health using machine learning algorithms. *J. Med. Syst.* **2018**, *42*, 88. [CrossRef] [PubMed]
15. Guo, C.; Tomson, G.; Keller, C.; Söderqvist, F. Prevalence and correlates of positive mental health in Chinese adolescents. *BMC Public Health* **2018**, *18*, 263. [CrossRef] [PubMed]
16. Witteveen, D.; Velthorst, E. Economic hardship and mental health complaints during COVID-19. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 27277–27284. [CrossRef]
17. Bubonya, M.; Cobb-Clark, D.A.; Wooden, M. Jobloss and the mental health of spouses and adolescent children. *IZAJ. LaborEcon.* **2017**, *6*, 6.
18. Brown, M.R.G. After the Fort McMurray wild fire there are significant increases in mental health symptoms ingrade 7–12 students compared to controls. *BMC Pyschiatry* **2019**, *19*, 18.
19. Pal, S.; Xu, T.; Yang, T.; Rajasekaran, S.; Bi, J. Hybrid-DCA: A double asynchronous approach for stochastic dual coordinate ascent. *J. Parallel Distrib. Comput.* **2020**, *143*, 47–66. [CrossRef]
20. Spiridonoff, A.; Olshevsky, A.; Paschalidis, I.C. Robust Asynchronous Stochastic Gradient-Push: Asymptotically Optimaland Network-Independent Performance for Strongly Convex Functions. *J. Mach. Learn. Res.* **2020**, *21*, 58.

21. Pu, S.; Olshevsky, A.; Paschalidis, I.C. A Sharp Estimate on the Transient Timeoff Distributed Stochastic Gradient Descent. *IEEE Trans. Automat. Contr.* **2022**, *67*, 5900–5915. [CrossRef]

22. Pu, S.; Olshevsky, A.; Paschalidis, I.C. Asymptotic Network Independence in Distributed Stochastic Optimization for Machine Learning. *IEEE Signal Process. Mag.* **2020**, *37*, 114–122. [CrossRef]

23. Mohsen, F.; Al-Saadi, B.; Abdi, N.; Khan, S.; Shah, Z. Artificial Intelligence-Based Methods for Precision Cardiovascular Medicine. *J. Pers. Med.* **2023**, *13*, 1268. [CrossRef]

24. Price, M.J. *Hello, C#! Welcome,. NET! In C# 8.0 and.NET Core 3.0—Modern Cross-Platform Development*, 4th ed.; Packt Publishing Ltd.: Birmingham, UK, 2019; pp. 1–69.

25. Perkins, B.; Hammer, J.V.; Reid, J.D. Introducing C#. In *Beginning C# 7 Programming with Visual Studio 2017*; Wiley: Hoboken, NJ, USA, 2018; pp. 3–13.

26. Shalev-Shwartz, S.; Tong, Z. Stochastic Dual Coordinate Ascent Methods for Regularized Loss Minimization. *arXiv* **2013**, arXiv:1209.1873.

27. Lu, X.; Yang, C.; Wu, Q.; Wang, J.; Wei, Y.; Zhang, L.; Li, D.; Zhao, L. Improved Reconstruction Algorithm of Wireless Sensor Network Based on BFGS Quasi-Newton Method. *Electronics* **2023**, *12*, 1267. [CrossRef]

28. Aggrawal, H.O.; Modersitzki, J. Hessian Initialization Strategies for L-BFGS Solving Non-linear Inverse Problems. *arXiv* **2021**, arXiv:2103.10010.

29. Asl, A.; Overton, M.L. Behavior of limited memory BFGS when applied to nonsmooth functions and their nesterov smoothings. *arXiv* **2020**, arXiv:2006.11336.

30. Bousbaa, Z.; Sanchez-Medina, J.; Bencharef, O. Financial Time Series Forecasting: A Data Stream Mining-Based System. *Electronics* **2023**, *12*, 2039. [CrossRef]

31. Benczúr, A.A.; Kocsis, L.; Pálovics, R. Online Machine Learning in Big Data Streams. *arXiv* **2018**, arXiv:1802.05872.

32. Ilboudo, W.E.L.; Kobayashi, T.; Sugimoto, K. Robust stochastic gradient descent with student-t distribution basedfirst-order momentum. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *33*, 1324–1337. [CrossRef]

33. Figalová, N.; Charvat, M. The Perceived Stress Scale: Reliability and validity study in the Czech Republic. *Ceskoslovenská Psychol.* **2021**, *65*, 46–59. [CrossRef]

34. Prasetya, A.; Purnama, D.; Prasetyo, F. Validity and Reliability of The Perceived Stress Scale with RASCH Model. *PSIKOPEDA-GOGIA J. Bimbing. Konseling* **2020**, *8*, 48–51. [CrossRef]

35. Maslach, C.; Jackson, S.E. The measurement of experienced burnout. *J. Occup. Behav.* **1981**, *2*, 99–113. [CrossRef]

36. Schaufeli, W.B.; Bakker, A.B.; Hoogduin, K.; Kladler, A.; Schaap, C. On the clinical validity of the Maslach Burnout Inventory and the Burnout Measure. *Psychol. Health* **2001**, *16*, 565–582. [CrossRef] [PubMed]

37. Checa, I.; Perales, J.; Espejo, B. Measurement in variance of the Satisfaction with Life Scale by gender, age, marital status and educational level. *Qual. Life Res. Int. J. Qual. Life Asp. Treat. Care Rehabil.* **2019**, *28*, 963–968. [CrossRef]

38. Diener, E.; Emmons, R.A.; Larsen, R.J.; Griffin, S. The Satisfaction with Life Scale. *J. Personal. Assess.* **1985**, *49*, 71–75. [CrossRef]

39. Prokopowicz, P.; Mikołajewski, D.; Mikołajewska, E. Intelligent System for Detecting Deterioration of Life Satisfaction as Tool for Remote Mental-Health Monitoring. *Sensors* **2022**, *22*, 9214. [CrossRef]

40. Rojek, I. Neural networks as prediction models for water intake in water supply system. In *Artificial Intelligence and Soft Computing—ICAISC 2008. Lecture Notes in Computer Science, 5097*; Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M., Eds.; Springer: Berlin/Heidelberg, Gemany, 2008; pp. 1109–1119. Available online: https://link.springer.com/chapter/10.1007/978-3-540-69731-2_104 (accessed on 31 August 2023).

41. Spoor, J.M.; Weber, J. Evaluation of process planning in manufacturing by a neural network based on an energy definition of Hopfield nets. *J. Intell. Manuf.* **2023**, 1–19. [CrossRef]

42. Teixeira, I.; Morais, R.; Sousa, J.J.; Cunha, A. Deep Learning Models for the Classification of Cropsin Aerial Imagery: A Review. *Agriculture* **2023**, *13*, 965. [CrossRef]

43. Rojek, I.; Mikołajewski, D.; Macko, M.; Szczepański, Z.; Dostatni, E. Optimization of Extrusion-Based 3D Printing Process Using Neural Networks for Sustainable Development. *Materials* **2021**, *14*, 2737. [CrossRef]

44. Rojek, I.; Mikołajewski, D.; Kotlarz, P.; Macko, M.; Kopowski, J. Intelligent system supporting technological process planning for machining and 3D printing. *Bull. Pol. Acad. Sci. Tech. Sci.* **2021**, *69*, e136722.

45. Mohammadi, E.K.; Talaie, H.R.; Azizi, M. A healthcare service quality assessment model usinga fuzzy best–worst method with application to hospitals within-patient services. *Healthc. Anal.* **2023**, *4*, 100241. [CrossRef]

46. Gajos, A.; Wójcik, G.M. Independent component analysis of EEG data for EGI system. *Bio-Algorithms Med-Syst.* **2016**, *12*, 67–72. [CrossRef]

47. Kawala-Janik, A.; Podpora, M.; Pelc, M.; Piatek, P.; Baranowski, J. Implementation of an inexpensive EEG headset for the pattern recognition purpose. In Proceedings of the 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), Berlin, Germany, 12–14 September 2013; Volume 1, pp. 399–403.