Article

Development of a Suicide Prediction Model for the Elderly Using Health Screening Data

Seo-Eun Cho [1], Zong Woo Geem [2] and Kyoung-Sae Na [1,*]

[1] Department of Psychiatry, Gachon University College of Medicine, Gil Medical Center, Incheon 21565, Korea; arztin01@gilhospital.com
[2] College of IT Convergence, Gachon University, Seongnam 13120, Korea; zwgeem@gmail.com
* Correspondence: ksna13@gmail.com; Tel.: +82-32-468-9932

Abstract: Suicide poses a serious problem globally, especially among the elderly population. To tackle the issue, this study aimed to develop a model for predicting suicide by using machine learning based on the elderly population. To obtain a large sample, the study used the big data health screening cohort provided by the National Health Insurance Sharing Service. By applying a machine learning technique, a predictive model that comprehensively utilized various factors was developed to select the elderly aged > 65 years at risk of suicide. A total of 48,047 subjects were included in the analysis. Individuals who died by suicide were older, and the number of men was significantly greater. The suicide group had a more prominent history of depression, with the use of medicaments significantly higher. Specifically, the prescription of benzodiazepines alone was associated with a high suicide risk. Furthermore, body mass index, waist circumference, total cholesterol, and low-density lipoprotein level were lower in the suicide group. We developed a model for predicting suicide by using machine learning based on the elderly population. This suicide prediction model can satisfy the performance to some extent by employing only the medical service usage behavior without subjective reports.

Keywords: suicide; the elderly; health screening cohort; machine learning; mental health

1. Introduction

Suicide is a serious social and cultural problem in modern society. According to the World Health Organization (WHO), over 700,000 people die by suicide every year [1]. Thus, the goal of the WHO Mental Health Action Plan 2013–2030 was to reduce the suicide rate in all WHO member states by one-third before 2030 [2].

Suicide among the elderly has emerged as a serious health and social problem in both the East and the West. The age group with the highest suicide rate in the United States in 2019 included individuals aged > 85 years, of which 21.12 of every 100,000 people died by suicide [3]. Furthermore, this group is the only age group for which the suicide rate has increased, considering the 2018 rate (19.07). The suicide rate for the elderly is also high in South Korea. Moreover, Korea has one of the highest suicide rates among the members of the Organization for Economic Co-operation and Development. According to the most recent data, the average suicide rate in OECD countries is 11.2, and South Korea ranked first with 23.0, followed by former Soviet states such as Lithuania (22.2) and Latvia (18.1). The country with the second highest suicide rate in Asia is Japan at 14.9 [4]. In 2019, the suicide rate was 46.6 per 100,000 people aged $\geq$ 65 years in South Korea, which is more than twice the average suicide rate and the highest among all age groups [5].

Older people are prone to loneliness and economic poverty. Unlike young adults and middle-aged people, they encounter difficulties forming active interpersonal relationships because people close to them continue to die or are already dead, and they experience repeated loss. Moreover, they become more vulnerable to economic factors as economic activities become difficult. In addition, the quality of life is reduced due to various chronic diseases, such as high blood pressure, diabetes, arthritis, dysuria, and cancer, and the

elderly feel greater physical discomfort. Usually, when they feel that the pain they are experiencing is intolerable, unchanging, and endless, they consider suicide as a way out [6]. From this perspective, factors such as interpersonal relationships, economic difficulties, physical illness, and functional decline all contribute to an increasing suicide rate. In a previous study on the elderly in Korea, the important factors leading to suicide attempts in suicidal ideation were male sex, education level, and age discrimination [7].

As the problem of suicide among the elderly is widespread and serious, prevention and preemptive interventions are required. In previous studies [8,9], suicide prevention interventions in the elderly population were found to reduce suicidal thoughts and suicide rates. For this reason, interest in the elderly suicide model has increased, including the development of protocols for models that describe and predict suicide in the elderly [10]. Thus, one of the common methods is to use a measurement tool. However, no measure has yet been proven to be effective in predicting future suicide. In fact, the United Kingdom's National Institute for Health and Care Excellence guideline recommends not to use risk assessment instruments to predict future suicide, suicide attempts, or self-harm [11]. For the elderly, expressing feelings in a way that is easy for others to understand can be also difficult, like in cases of masked depression [12]. In addition, several suicide measures include many questions. There are many elders who cannot use such measures as smoothly as younger adults, due to deterioration of vision and cognitive ability because of aging. Thus, models predicting the suicide risk among the elderly should fully consider this.

Due to the fact that suicide is a very rare event, an approach taking this into account is required. In some cases, groups for which suicide is more common than for the general population, such as individuals suffering from depression or bipolar disorder, are selected for studies [13]. In other cases, a suicide attempt instead of suicide is used as a key outcome indicator [14]. However, both approaches have limitations. Limiting subjects to a clinical population only prevents the inclusion of subjects who do not fit into the therapeutic system. Furthermore, although a suicide attempt is the single largest risk factor for suicide, and is most similar to suicide, suicide attempts and suicide have distinct characteristics [15,16].

Accordingly, the best method is to predict suicide in the general population of a community. A previous study on the suicide prediction model [17], which included 2,960,929 subjects from 2009 to 2015, was limited in that only outpatients from the department of psychiatric medicine were included. In a systematic review [18] of studies that included 64 unique suicide prediction models, the general population was not included. The predictive validity of suicide was very low in most studies and practical utility was limited. As is widely known, a study of community-dwelling suicidal ideators over 65 years of age found that both depression and stress predict the intensity of suicidal thoughts [19]. There has also been a study [20] showing that both big data and machine learning methods need to be improved for accurate suicide prediction.

The biggest challenge in predicting suicide in the general population is the need to obtain a large sample. For this, the authors used a big data health screening dataset provided by the National Health Insurance Sharing Service (NHISS) [21]. In addition, in a socio-cultural environment with a high social prejudice against suicide, such as in Korea, there is a limit to suicide prevention because predictions and interventions are limited to groups at high risk of suicide. Therefore, using health check-up data that are generally available for the entire population, it is possible to reduce problems such as avoidance due to stigma. By applying a machine-learning technique, a predictive model that comprehensively utilizes various factors was developed to select the elderly with a high risk of suicide.

2. Materials and Methods

2.1. Data and Subjects

We used the health screening cohort database (DB) provided by the NHISS. Ten percent of the 5.15 million people eligible for health insurance aged between 40 and 79 years as of the end of December 2002, among those who undertook general health screening between

2002 and 2003, were selected by a simple random sampling technique. Qualifications and income information (socioeconomic variables), hospital and clinic usage history, health screening results, and healthcare institution information from 2002 to 2015 (14 years) for 514,866 people were constructed in a cohort format. These data were constructed anonymously in a cohort format so that these data can be used for research purposes in South Korea [21].

We then specifically targeted patients aged $\geq$ 65 years who undertook health screening in 2009 from the health screening cohort DB, and this was done for several reasons. Firstly, various measurement variables and questionnaire items have changed since 2008 and 2009, resulting in inconsistency. Secondly, since the subjects were older than 65, more people die over time, and thus, the subjects at the baseline of the analysis inevitably become scarcer. Accordingly, those who took health screening in 2009, which was the earliest possible period, were targeted.

This study was approved by the institutional review board (IRB No. GDIRB2019-253) of the Gil Medical Center.

2.2. Variables

We used the following variables in the study: age, sex, disability, medical benefits, diagnosis of depressive disorder in the past year as of the date of examination, psychiatric drug use in the past year (antidepressants, mood stabilizers, sleeping pills, benzodiazepines, anti-psychotic drugs, and cognitive enhancement drugs), ranking of insurance premium payment at the time of examination, body mass index (BMI), waist circumference, systolic blood pressure, fasting plasma glucose, total cholesterol, triglyceride, creatinine, hemoglobin, aspartate transaminase, and gamma-glutamyl transpeptidase. All psychiatric diagnoses were in accordance with the *International Statistical Classification of Diseases and Related Health Problems 10th Revision* (ICD-10) [22]. When the diagnosis fell under F32 and F33, it was defined as a depressive disorder. Antidepressants included citalopram, escitalopram, paroxetine, fluoxetine, fluvoxamine, venlafaxine, desvenlafaxine, duloxetine, milnacipran, bupropion, mirtazapine, tianeptine, trazodone, clomipramine, amitriptyline, imipramine, amoxapine, and moclobemide. Mood stabilizers included lithium, lamotrigine, valproic acid (or divalproex), and carbamazepine. Antipsychotic drugs included chlorprothixene, haloperidol, levomepromazine, molindone, perphenazine, pimozide, sulpiride, amisulpride, aripiprazole, blonanserin, clozapine, olanzapine, quetiapine, risperidone, ziprasidone, and zotepine. Sleeping pills included zolpidem and doxepine. Benzodiazepines included flunitrazepam, flurazepam, chlordiazepoxide, clobazam, triazolam, clonazepam, diazepam, lorazepam, mexazolam, bromazepam, clotiazepam, alprazolam, and etizolam. Cognitive enhancers included donepezil, galantamine, rivastigmine, and memantine.

According to the common recommendations of academia [23] and various studies [17,24,25], cases falling under the ICD-10 codes X60–X84 or Y10–Y34 were defined as suicide. The tracking period for suicide was until 2015.

2.3. Data Preprocessing and Machine Learning

The extraction of data necessary for research from the entire health screening cohort DB was carried out using the SAS Enterprise Guide 7.1 (SAS Institute Inc., Cary, NC, USA). All cases with missing values were excluded from the analysis. All subsequent work was carried out by using R 3.3.0 [26] and RStudio 1.0.136 [27]. In addition, all data were divided into a training set (0.7) and a test set (0.3) at a ratio of 7:3, respectively. Considering that suicide is very rare, the synthetic minority over-sampling technique was used [28]. No structural transformation of the data took place in the test set. The entire training dataset was trained within the training set through 10-fold cross-validation. The area under the receiver operating characteristic curve (AUC) was established as the main performance indicator. In addition, an overall accuracy, sensitivity, specificity, negative predictive value, and positive predictive value were calculated. Performance measurements, such as hyper-

parameter tuning, confusion matrix composition, and AUC, were all performed using the *caret* library.

For the algorithm, random forest was used. In general, the decision tree model has the advantage that variable importance and classification mechanisms can be intuitively known and the computing cost is low. However, the decision tree is rarely used as an algorithm for solving problems in practice since it has the crucial problem of being vulnerable to overfitting. Random forest is the ensemble model to compensate for the shortcomings of this decision tree [29]. It maximizes predictive power when new data (i.e., test set) are given rather than used for training by minimizing the overfitting of decision tree. Thanks to this advantage, random forest has been used as a practical problem-solving algorithm in various fields [30,31].

To find out which variables contributed significantly to optimization of the predictive model, we calculated variable importance by the principle of "mean decreases in accuracy" [32]. The process of estimating the importance of variable j is as follows: Several cases were not sampled and were called out-of-bag because bagging permits cases to be sampled more than once for the same classifier. The out-of-bag samples hand down to the tree, and the prediction accuracy then decreases when training by utilizing one of the trees is performed. Subsequently, the variable j was changed in the out-of-bag samples at random, and the accuracy was calculated again. Before and after the permutation over all trees, the raw score for the importance of variable j was calculated by averaging the gap in out-of-bag errors. Subsequently, the score was returned to normal by utilizing the standardized deviation of the difference. Eventually, the score was reduced so that the minimum value was set to 0 and the maximum value was set to 100. Using care, a confusion matrix, hyperparameter tuning, and performance measurement were attained. In brief, the prediction performance of the out-of-bag part of the data was recorded for each tree. Then, the prediction performance was measured and recorded each time, while the predictors were permuted alternatively. Based on all trees, the average value of the difference between these two prediction performances was obtained and then normalized according to the standard error. The final prediction performance was measured by using test data, which were never used in the training phase. An overall overview of machine learning is presented in Figure 1.
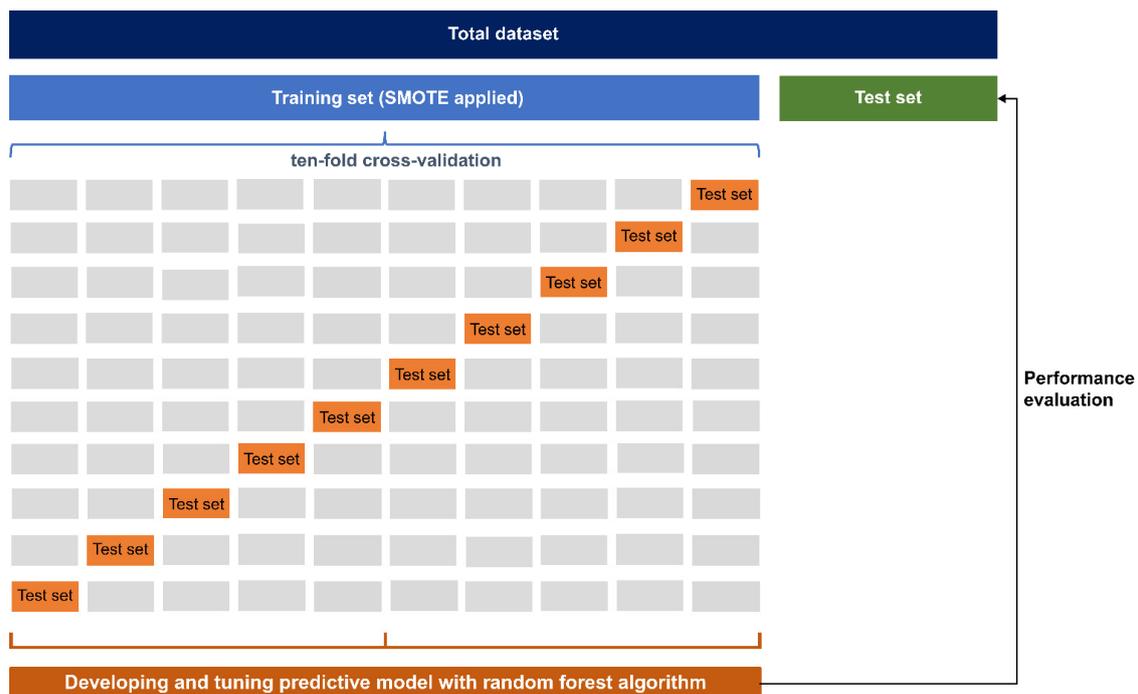

Figure 1. The overview of machine learning.

Int. J. Environ. Res. Public Health **2021**, *18*, 10150

5 of 10

3. Results

3.1. Socio-Demographic and Clinical Characteristics

A total of 48,047 subjects were included in the analysis. Among them, the number of people who died by suicide was 100 (0.21%). Their sociodemographic and clinical characteristics are presented in Table 1. Those who died by suicide were older than those who did not, and significantly more men than women died by suicide. Among those who underwent health screening, the percentage of men and women was almost equal, but in the suicide group, 72% were men. Moreover, the suicide group had a greater history of depression and significantly higher use of antidepressants, mood stabilizers, antipsychotic drugs, sleeping pills, and benzodiazepines than the non-suicide group. Regarding examination results, BMI, waist circumference, total cholesterol, and low-density lipoprotein (LDL) levels were lower in the suicide group.

Table 1. Comparison of sociodemographic and economic characteristics and medical use between suicide and non-suicide groups.

Variables	Non-Suicide (*n* = 47,947)	Suicide (*n* = 100)	t or χ^2	*p*-Value
Age, years	72.1 ± 4.41	74.1 ± 5.27	−3.72	0.0003 [a]
Sex, male	24,228 (50.53%)	72 (72%)	18.4	<0.0001 [b]
Disability, existence	1024 (2.14%)	7 (7%)	11.245	0.0008 [b]
Medical benefit, existence	189 (0.39%)	0 (0%)	0.393	0.529 [b]
Depression diagnosis, existence	2445 (5.1%)	17 (17%)	28.071	<0.0001 [b]
Taking psychiatric drugs, existence				
Antidepressants	4095 (8.54%)	23 (23%)	26.625	<0.0001 [b]
Mood stabilizer	576 (1.2%)	4 (4%)	0.011	0.011 [b]
Antipsychotics	472 (0.98%)	4 (4%)	0.002	0.002 [b]
Sleeping pills	1806 (3.77%)	16 (16%)	<0.0001	<0.0001 [b]
Benzodiazepines	13,755 (28.69%)	59 (59%)	44.761	<0.0001 [b]
Cognitive enhancer	412 (0.86%)	0 (0%)	0.8667	0.352 [b]
Body measurement				
BMI, kg/m2	23.8 ± 3.1	21.8 ± 2.94	6.8	<0.0001 [a]
Waist circumference, cm	83.17 ± 8.33	80.97 ± 8.21	2.67	0.009 [a]
Systolic blood pressure, mmHg	129.8 ± 16.01	128.3 ± 16.82	0.95	0.345 [a]
Diastolic blood pressure, mmHg	78.04 ± 10.02	77.89 ± 10.57	0.15	0.879 [a]
Blood biochemical results				
Fasting plasma glucose, mg/dL	102.6 ± 26.08	102.5 ± 2.11	0.05	0.958 [a]
Total cholesterol, mg/dL	196.4 ± 38.76	185.7 ± 40.36	2.65	0.009 [a]
Hemoglobin, g/dL	13.41 ± 1.45	13.27 ± 1.8	0.81	0.4225 [a]
AST, U/L	26.55 ± 15.7	32.17 ± 36.99	−1.52	0.132 [a]
ALT, U/L	22.63 ± 16.84	24.96 ± 24.61	−0.94	0.351 [a]
GGT, U/L	34.09 ± 49.93	50.36 ± 89.42	−1.82	0.072 [a]
TG, mg/dL	138.6 ± 82.15	131.2 ± 83.99	0.9	0.369 [a]
HDL, mg/dL	55.17 ± 35.98	59.47 ± 55.59	−0.77	0.441 [a]
LDL, mg/dL	116.1 ± 38.74	107.6 ± 39.08	2020	0.028 [a]
Creatinine, mg/dL	1.05 ± 1.11	1.21 ± 1.13	−1.42	0.203 [a]

Data are presented as the mean ± standard deviation or number (%). Abbreviations: SD, standard deviation; AST, aspartate transaminase; ALT, alanine transaminase; GGT, gamma-glutamyl transpeptidase; TG, triglyceride; HDL, high-density lipoprotein; LDL, low-density lipoprotein. Statistical tests were performed using Student's *t*-test [a] and the chi-square test [b].

3.2. Prediction Model Performance and Variable Importance

The AUC, which is the most preferred indicator of the prediction model, was 0.818 (Figure 2). In addition, overall accuracy was 0.832, sensitivity was 0.600, specificity was 0.833, negative predictive value was 0.999, and positive predictive value was 0.007. The variable with the highest importance was the history of taking benzodiazepines, followed by BMI, age, and history of taking sleeping pills (Figure 3).
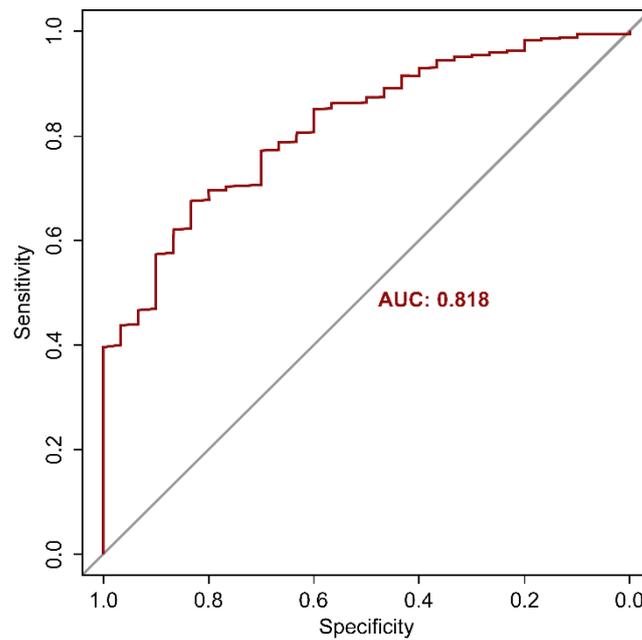
Int. J. Environ. Res. Public Health **2021**, *18*, 10150

6 of 10

Figure 2. Area under the receiver operating characteristic curve of the suicide prediction model.
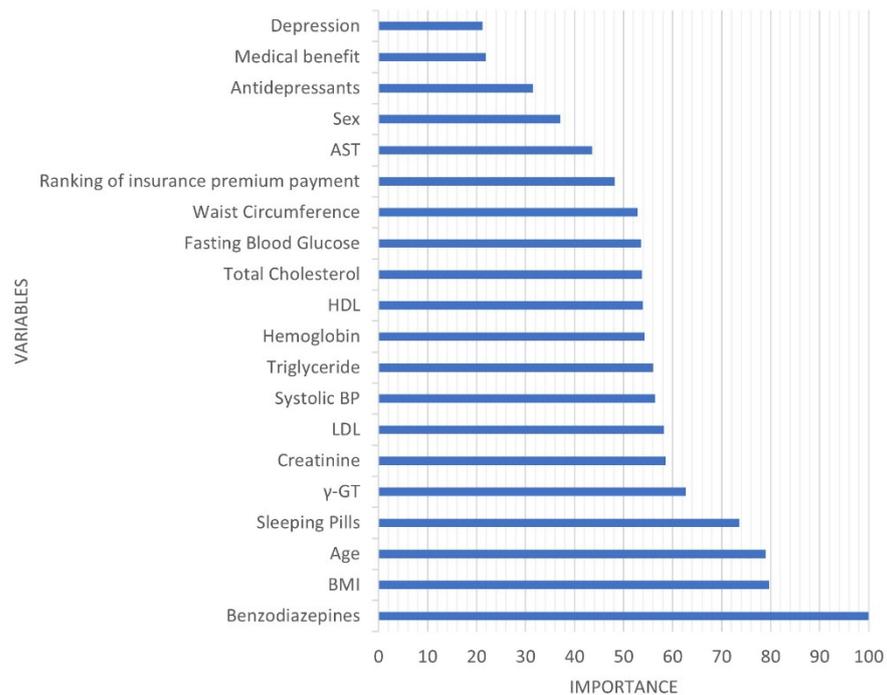

Figure 3. Variables' importance in the suicide prediction model.

4. Discussion

We developed a model to predict the high-risk group for suicide using machine learning algorithms for the elderly aged > 65 years who underwent health screening. The fact that benzodiazepines had the most importance in this study has great implications. Benzodiazepines had the highest statistic among all variables with the χ^2 value of 44.761 in a simple frequency analysis between the suicide and non-suicide groups. This is consistent with the results of previous studies, according to which prescription of benzodiazepines alone is associated with a high risk of suicide, even if not used as a means of suicide [33,34]. The mechanism by which the prescribed benzodiazepines and sleeping pills, collectively known as sedatives and hypnotics, are associated with suicide in elderly remains unclear.

First, a possibility of misusing the prescribed sedatives and hypnotics for suicidal purposes can be considered. However, substance misuse is not a factor highly associated with suicide deaths, although it is mainly used for suicide attempts [35]. In 2019, the means that had the greatest impact on the suicide rate of the elderly older than 65 in South Korea was hanging (51.0%), followed by falling (18.0%), and taking pesticides (16.1%). Drugs accounted for only 2.3% of all means of suicide [5]. Another possibility is that the elderly who were prescribed sedatives and hypnotics had greater anxiety, agitation, and tension. Unfortunately, it is difficult to directly determine whether the deceased elderly used prescribed sedatives and hypnotics as a means, and to what extent psychiatric symptoms were actually experienced, due to the design of this study.

Another characteristic of suicide victims is that they had lower weight and slimmer waistlines, and lower total cholesterol and LDL levels. Low BMI, waist circumference, and LDL are all associated with low total cholesterol levels. The results of studies which support the connection between low cholesterol and suicide have been reported for a long time [36,37]. Recent studies elucidated the genetic overlap between cholesterol metabolism and suicide [38]. Additionally, among the medical parameters, GGT, creatinine, and systolic blood pressure were of high importance. Although the t-test indicated no significant difference, the mean GGT and creatinine levels were higher in the suicide deaths than in the control group, and the systolic blood pressure was slightly lower. Higher GGT levels were associated with a higher risk of suicide as alcohol intake increased. Creatinine is more likely to appear higher as the kidney function is reduced, drugs are taken, or with older age. Furthermore, the higher the drug intake, the more severe the underlying disease may be. Stress and suicide rates have been consistently found to be highly correlated [39]. As previously reported [40], stress may cause strain on the kidneys. It has been found that low systolic blood pressure is associated with suicidal ideation in the general Korean population [41], which is consistent with the results of this study. If the systolic blood pressure is low, there may be a possibility of the perception that overall negative emotions and pain cannot be suppressed [42].

Regarding the performance of the prediction model, it had the performance of AUC 0.818 and accuracy 0.832. The AUC value criterion for excellence has not been clearly established. A satisfactory AUC value criterion may vary due to various factors, such as data distribution, research area, and design. Generally, a value > 0.7 is considered acceptable, a value $\geq$ 0.8 is good, and a value > 0.9 is excellent [43]. This study was designed as a retrospective cohort follow-up approach. Another study compared the effect sizes from several studies and conducted follow-up analysis. It stated that an AUC score of 0.714 corresponds to Cohen's d of 0.8 [44]. A Cohen's d > 0.8 is considered to indicate a large effect size [45]. However, the low positive predictive value is considered a task to be overcome in the future. The probability of suicide in the general population is very low, at about 0.00025. For this reason, a low positive predictive value has always posed a problem in studies predicting suicide [46]. A review reported that positive predictive values of measures evaluating suicide risk in people receiving primary healthcare ranged from 0.01 to 0.2 [47]. In another meta-analysis, the positive predictive values of the tool, which predicts suicide risk, were analyzed for psychiatric patients only [48]. As a result, it was found to be 0.055 (95% confidence interval [CI] 0.035–0.085) in the high-risk and 0.09 (95% CI 0.005–0.017) in the low-risk group. Although the positive predictive value obtained in this study is insufficient as a par value, it is 28 times higher than the suicide rate of the general population.

There are several limitations to this study. Firstly, we encountered a difficulty in elucidating the mechanism of the prediction model due to the design of the retrospective cohort study. According to ICD-10, detailed methods of suicide are to be classified as sensitive information and are not provided to researchers. Thus, the role of the prescribed drug was not fully examined. Secondly, clinical information highly related to suicide, such as the severity of psychiatric symptoms, was unavailable. Thirdly, the study was not free from the selection bias, which fundamentally excluded people who had not undergone

health screening. For small-subject studies, prospective randomized controlled studies and study designs that include more clinical information highly relevant to suicide may be possible. However, there is also a limitation in that it is difficult to design a study as a result of suicide death, and since the number of participants is small, there may be selective bias and low reliability.

Based on the results obtained while compensating for these limitations, practical data can be provided for developing suicide prevention programs and seeking counseling approaches to effectively prevent suicide from the stage of suicidal thoughts. The suicide prediction model built using big data can be widely used in most areas, such as national health checkups, emergency-room-based follow-up management projects for suicide attempters, and outpatient and inpatient treatment at medical institutions. The effectiveness of suicide prevention projects can be increased by enabling additional interest and intervention in elderly suicide attempters, especially those with higher actual suicide risk, using machine learning algorithms. Additional studies for tailored approaches to individual subjects should be conducted in the future.

5. Conclusions

This study developed a model for predicting suicide that occurs infrequently by using machine learning based on the elderly population aged > 65 years from the community. This suicide prediction model can satisfy the performance to some extent by employing only the medical service usage behavior without subjective reports from the elderly, although the model has many aspects that need to be improved in terms of a positive predictive value. By continuing research in this area and addressing the limitations of this study, it will be possible to implement the suicide prediction model practically in real life. Additionally, in suicide research using big data machine learning methods, structural suicide risk assessment tools should be used, and both the assessment of suicide risk and analysis to predict the most appropriate treatment options for individual patients should be performed.

References

1. World Health Organization. Suicide. Available online: https://www.who.int/news-room/fact-sheets/detail/suicide (accessed on 28 July 2021).
2. World Health Organization. WHO Mental Health Action Plan 2013–2030. Available online: https://www.who.int/publications/i/item/9789241506021 (accessed on 28 July 2021).
3. American Foundation for Suicide Prevention. Suicide statistics. Available online: https://afsp.org/suicide-statistics/#:~{}:text=In%202019%2C%20the%20suicide%20rates,2018%20to%202020.12%20in%202019 (accessed on 28 July 2021).
4. OECD. Suicide Rates (Indicator). Available online: https://data.oecd.org/healthstat/suicide-rates.htm (accessed on 25 July 2021).
5. Ministry of Health and Welfare; Korea Foundation for Suicide Prevention. *2021 White Paper on Suicide Prevention*; Seoul, Korea, 2021.
6. Chiles, J.A.; Strosahl, K.D.; Roberts, L.W. *Clinical Manual for the Assessment and Treatment of Suicidal Patients*, 2nd ed.; American Psychiatric Association Publishing: Washington, DC, USA, 2019.
7. Ko, Y.; Han, S.Y.; Jang, H.Y. Factors Influencing Suicidal Ideation and Attempts among Older Korean Adults: Focusing on Age Discrimination and Neglect. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1852. [CrossRef]
8. Lapierre, S.; Erlangsen, A.; Waern, M.; De Leo, D.; Oyama, H.; Scocco, P.; Gallo, J.; Szanto, K.; Conwell, Y.; Draper, B.; et al. A systematic review of elderly suicide prevention programs. *Crisis* **2011**, *32*, 88–98. [CrossRef]
9. Zeppegno, P.; Gattoni, E.; Mastrangelo, M.; Gramaglia, C.; Sarchiapone, M. Psychosocial Suicide Prevention Interventions in the Elderly: A Mini-Review of the Literature. *Front. Psychol.* **2018**, *9*, 2713. [CrossRef]
10. Rostami, M.; Younesi, S.J.; Mohammadi Shahboulaghi, F.; Malakouti, S.K.; Foroughan, M. Models of suicide in elderly: A protocol for a systematic review. *BMJ Open* **2018**, *8*, e022087. [CrossRef]
11. NICE. *Self-Harm in Over 8s: Long-Term Management*; NICE: London, UK, 2011.
12. Shetty, P.; Mane, A.; Fulmali, S.; Uchit, G. Understanding masked depression: A Clinical scenario. *Indian J. Psychiatry* **2018**, *60*, 97–102. [CrossRef] [PubMed]
13. Monson, E.T.; Shabalin, A.A.; Docherty, A.R.; DiBlasi, E.; Bakian, A.V.; Li, Q.S.; Gray, D.; Keeshin, B.; Crowell, S.E.; Mullins, N.; et al. Assessment of suicide attempt and death in bipolar affective disorder: A combined clinical and genetic approach. *Transl. Psychiatry* **2021**, *11*, 379. [CrossRef] [PubMed]
14. Naifeh, J.A.; Ursano, R.J.; Stein, M.B.; Mash, H.B.H.; Aliaga, P.A.; Fullerton, C.S.; Dinh, H.M.; Vance, M.C.; Wynn, G.H.; Kao, T.C.; et al. Risk of suicide attempt in reserve versus active component soldiers during deployment to the wars in Iraq and Afghanistan. *Suicide Life-Threat. Behav.* **2021**. [CrossRef] [PubMed]
15. Barry, R.; Rehm, J.; de Oliveira, C.; Gozdyra, P.; Kurdyak, P. Rurality and Risk of Suicide Attempts and Death by Suicide among People Living in Four English-speaking High-income Countries: A Systematic Review and Meta-analysis. *Can. J. Psychiatry* **2020**, *65*, 441–447. [CrossRef] [PubMed]
16. Jordan, J.T.; McNiel, D.E. Characteristics of persons who die on their first suicide attempt: Results from the National Violent Death Reporting System. *Psychol. Med.* **2020**, *50*, 1390–1397. [CrossRef] [PubMed]
17. Simon, G.E.; Johnson, E.; Lawrence, J.M.; Rossom, R.C.; Ahmedani, B.; Lynch, F.L.; Beck, A.; Waitzfelder, B.; Ziebell, R.; Penfold, R.B.; et al. Predicting Suicide Attempts and Suicide Deaths Following Outpatient Visits Using Electronic Health Records. *Am. J. Psychiatry* **2018**, *175*, 951–960. [CrossRef] [PubMed]
18. Belsher, B.E.; Smolenski, D.J.; Pruitt, L.D.; Bush, N.E.; Beech, E.H.; Workman, D.E.; Morgan, R.L.; Evatt, D.P.; Tucker, J.; Skopp, N.A. Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation. *JAMA Psychiatry* **2019**, *76*, 642–651. [CrossRef]
19. Park, C.H.K.; Lee, J.W.; Lee, S.Y.; Shim, S.H.; Kim, S.G.; Lee, J.; Kim, M.H.; Paik, J.W.; Cho, S.J.; Moon, J.J.; et al. Characteristics of the "young-old" and "old-old" community-dwelling suicidal Ideators: A longitudinal 6-month follow-up study. *Compr. Psychiatry* **2019**, *89*, 67–77. [CrossRef] [PubMed]
20. Kessler, R.C.; Bernecker, S.L.; Bossarte, R.M.; Luedtke, A.R.; McCarthy, J.F.; Nock, M.K.; Pigeon, W.R.; Petukhova, M.V.; Sadikova, E.; VanderWeele, T.J.; et al. The Role of Big Data Analytics in Predicting Suicide. In *Personalized Psychiatry*; Springer: New York, NY, USA, 2019; pp. 77–98. [CrossRef]
21. National Health Insurance Service. National Health Insurance Sharing Service. Available online: https://nhiss.nhis.or.kr/bd/ab/bdaba000eng.do (accessed on 15 July 2021).
22. World Health Organization. *ICD-10: International Statistical Classification of Diseases and Related Health Problems: Tenth Revision*, 2nd ed.; World Health Organization: Geneva, Switzerland, 2004.
23. Bakst, S.S.; Braun, T.; Zucker, I.; Amitai, Z.; Shohat, T. The accuracy of suicide statistics: Are true suicide deaths misclassified? *Soc. Psychiatry Psychiatr. Epidemiol.* **2016**, *51*, 115–123. [CrossRef] [PubMed]
24. Fazel, S.; Wolf, A.; Larsson, H.; Mallett, S.; Fanshawe, T.R. The prediction of suicide in severe mental illness: Development and validation of a clinical prediction rule (OxMIS). *Transl. Psychiatry* **2019**, *9*, 98. [CrossRef] [PubMed]
25. Lindh, A.U.; Beckman, K.; Carlborg, A.; Waern, M.; Salander Renberg, E.; Dahlin, M.; Runeson, B. Predicting suicide: A comparison between clinical suicide risk assessment and the Suicide Intent Scale. *J. Affect. Disord.* **2020**, *263*, 445–449. [CrossRef] [PubMed]
26. R Core Team. *R: A Language and Environment for Statistical Computing.*; R Foundation for Statistical Computing: Vienna, Austria, 2016.

Int. J. Environ. Res. Public Health **2021**, *18*, 10150

10 of 10

27. RStudio Team. *RStudio: Integrated Development for R*; RStudio, Inc.: Boston, MA, USA, 2012.
28. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
29. Breiman, L. Random Forest. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
30. Erickson, J.; Abbott, K.; Susienka, L. Automatic address validation and health record review to identify homeless Social Security disability applicants. *J. Biomed. Inform.* **2018**, *82*, 41–46. [CrossRef]
31. Zhang, L.; Fabbri, D.; Lasko, T.A.; Ehrenfeld, J.M.; Wanderer, J.P. A System for Automated Determination of Perioperative Patient Acuity. *J. Med. Syst.* **2018**, *42*, 123. [CrossRef]
32. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: New York, NY, USA, 2017; pp. 593–594. [CrossRef]
33. Cato, V.; Hollandare, F.; Nordenskjold, A.; Sellin, T. Association between benzodiazepines and suicide risk: A matched case-control study. *BMC Psychiatry* **2019**, *19*, 317. [CrossRef]
34. Dodds, T.J. Prescribed Benzodiazepines and Suicide Risk: A Review of the Literature. *Prim. Care Companion CNS Disord.* **2017**, *19*. [CrossRef] [PubMed]
35. Han, D.G.; Kang, S.G.; Cho, S.J.; Cho, S.E.; Na, K.S. Suicide Methods According to Age and Sex: An Analysis of Data of 239,565 Suicide Victims in the Republic of Korea From 1991 to 2015. *J. Nerv. Ment. Dis.* **2018**, *206*, 770–775. [CrossRef] [PubMed]
36. Messaoud, A.; Mensi, R.; Mrad, A.; Mhalla, A.; Azizi, I.; Amemou, B.; Trabelsi, I.; Grissa, M.H.; Salem, N.H.; Chadly, A.; et al. Is low total cholesterol levels associated with suicide attempt in depressive patients? *Ann. Gen. Psychiatry* **2017**, *16*, 20. [CrossRef] [PubMed]
37. Asellus, P.; Nordstrom, P.; Jokinen, J. Cholesterol and CSF 5-HIAA in attempted suicide. *J. Affect. Disord.* **2010**, *125*, 388–392. [CrossRef]
38. Knowles, E.E.M.; Curran, J.E.; Meikle, P.J.; Huynh, K.; Mathias, S.R.; Goring, H.H.H.; VandeBerg, J.L.; Mahaney, M.C.; Jalbrzikowski, M.; Mosior, M.K.; et al. Disentangling the genetic overlap between cholesterol and suicide risk. *Neuropsychopharmacology* **2018**, *43*, 2556–2563. [CrossRef]
39. Niculescu, A.B.; Levey, D.F.; Phalen, P.L.; Le-Niculescu, H.; Dainton, H.D.; Jain, N.; Belanger, E.; James, A.; George, S.; Weber, H.; et al. Understanding and predicting suicidality using a combined genomic and clinical risk assessment approach. *Mol. Psychiatry* **2015**, *20*, 1266–1285. [CrossRef]
40. Su, G.; Song, H.; Lanka, V.; Liu, X.; Fang, F.; Valdimarsdottir, U.A.; Carrero, J.J. Stress Related Disorders and the Risk of Kidney Disease. *Kidney Int. Rep.* **2021**, *6*, 706–715. [CrossRef]
41. Joung, K.I.; Cho, S.I. Association of low blood pressure with suicidal ideation: A cross-sectional study of 10,708 adults with normal or low blood pressure in Korea. *BMC Public Health* **2018**, *18*, 200. [CrossRef]
42. Delgado, L.C.; Vila, J.; Reyes del Paso, G.A. Proneness to worry is negatively associated with blood pressure and baroreflex sensitivity: Further evidence of the blood pressure emotional dampening hypothesis. *Biol. Psychol.* **2014**, *96*, 20–27. [CrossRef]
43. Hosmer, D.W.; Lemeshow, S. Assessing the Fit of the Model. In *Applied Logistic Regression*; Shewhart, W.A., Wilks, S.S., Hosmer, D.W., Lemeshow, S., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2000; Chapter 5; pp. 143–202. [CrossRef]
44. Rice, M.E.; Harris, G.T. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law Hum. Behav.* **2005**, *29*, 615–620. [CrossRef]
45. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Erlbaum: Hillsdale, NJ, USA, 1988.
46. Kessler, R.C.; Bossarte, R.M.; Luedtke, A.; Zaslavsky, A.M.; Zubizarreta, J.R. Suicide prediction models: A critical review of recent research with recommendations for the way forward. *Mol. Psychiatry* **2020**, *25*, 168–179. [CrossRef]
47. Abarca, C.; Gheza, C.; Coda, C.; Elicer, B. Literature review to identify standardized scales for assessing adult suicide risk in the primary health care setting. *Medwave* **2018**, *18*, e7246. [CrossRef]
48. Large, M.; Kaneson, M.; Myles, N.; Myles, H.; Gunaratne, P.; Ryan, C. Meta-Analysis of Longitudinal Cohort Studies of Suicide Risk Assessment among Psychiatric Patients: Heterogeneity in Results and Lack of Improvement over Time. *PLoS ONE* **2016**, *11*, e0156322. [CrossRef] [PubMed]