Fitness-To-Drive Assessment of Older Drivers Based on Multi-Classification Predictive Models

Peer-reviewed author version

Yue, Xinyi; BAO, Qiong; SHEN, Yongjun; Zhou, Muxiong & WETS, Geert (2024) Fitness-To-Drive Assessment of Older Drivers Based on Multi-Classification Predictive Models. In: Wang, W.; Lu, G. ; Si, Y. (Ed.) Smart transportation and green mobility safety, GITSS 2022, SPRINGER-VERLAG SINGAPORE PTE LTD, p. 271 -281.

DOI: 10.1007/978-981-97-3005-6_19
Handle: http://hdl.handle.net/1942/45876

Fitness-to-drive assessment of older drivers based on multi-classification predictive models

Xinyi Yue[1], Qiong Bao[1], Yongjun Shen[1,*], Muxiong Zhou[2], Geert Wets[3]
[1]School of Transportation, Southeast University, Nanjing, 210096, China
[2]Traffic Management Research Institute of the Ministry of Public Security, Wuxi, 214151, China
[3]Transportation Research Institute (IMOB), Hasselt University, 3500 Hasselt, Belgium
*Correspondence: shenyongjun@seu.edu.cn

Abstract: Aging has become a global issue, which is accompanied with the increase of older drivers on the road. To reduce the road safety problems caused by older drivers, it is necessary to assess their fitness-to-drive. In this study, by identifying three categories of older drivers from on-road driving test (i.e., “fit to drive”, “undetermined” and “unfit to drive”), six indicators from their functional ability tests and the simulated driving test were extracted, and three machine learning methods, i.e., decision tree, random forest and support vector machine, were applied to conduct the fitness-to-drive assessment among a number of older drivers. The result showed that the support vector machine achieved the best performance, with the highest accuracy rate over 80%, which implies the feasibility of using the proposed assessment procedure as an alternative way to the on-road test for the fitness-to-drive assessment for older drivers.

Keywords: older drivers, fitness-to-drive, multi-classification, support vector machine

1 Introduction

With the improvement of medical treatment and life quality, aging has become an inevitable problem all over the world. According to the results of the seventh Chinese census in 2020, the number of people aged 65 and above reached 191 million, accounting for 13.5% of the country's total population. With the deepening of the aging trend, the number of older drivers who intend to drive is continuing to increase as well. However, those unfit-to-drive ones are prone to cause traffic safety problems, it is therefore necessary for older drivers to conduct regular and effective fitness-to-drive assessment.

Road test is the most commonly used fitness-to-drive evaluation method in present, this method relies on a professional tester to rate whether an older driver is fit to drive or not. The evaluation process of the road test is intuitive, therefore it is regarded as “the gold standard” [1-2]. However, the road test requires a lot of manpower, financial resources and time [3]. Meanwhile, there is risk for older drivers who are no longer fit to drive but do not know about it, it is therefore urgent to develop an alternative and safer way to conduct fitness-to-drive assessment for older drivers. At least it should be an effective supplementation to the on-road driving test.

The languishing of the physical and mental function of the elderly will lead to the deterioration of their driving function. Diminution of vision and impaired visual field are the most common sensory problems affecting the driving performance of older drivers. The elderly generally have low contrast sensitivity and are more susceptible to glare. Eye diseases such as cataracts, glaucoma will cause more severe visual problems [5-6]. Balance, flexibility, endurance and reflex ability will deteriorate with age as well, leading to a decline in operation skills when driving [7]. In addition, there are some older drivers who have cognitive impairment. To cope with these problems, fitness-to-drive assessment of the elderly can be conducted by combining driving ability test with visual function, physical ability and cognitive ability tests in hospital. Typical tests include ADReS [8], DriveAble [9], DriveSafe/DriveAware [10] and SMCTests [11]. However, these test groups have high requirements on the professionalism of the doctors and lack of pervasiveness, the evaluation results of some test groups such as ADReS are controversial to a certain extent [12].

The emergence of simulated driving provides a channel for low-risk driving tests in controlled scenarios. Compared to the road test, simulated driving costs less and can reduce the driving pressure of participants. Moreover, traffic flow, weather conditions and dangerous events in the scene, which cannot be controlled in the road test, can be set according to practical needs. The achievements of simulated driving have been recognized in some researches and can be an alternative to on-road test.

Fitness-to-drive assessment can be treated as a classification issue. Logistic regression is traditionally used to solve classification problems but proved to be flawed, the development of machine learning provides new ideas for classification prediction. Commonly used machine learning algorithms are k-nearest neighbors (KNN), bayes Theorem, decision tree (DT), random forest (RF), support vector machine (SVM) and gradient boosted machine (GBM). In recent years, some of these models have been used to assess fitness-to-drive [13-14].

Piersma [1] found that when clinical interview, neuropsychological assessment and simulated driving assessment are combined, fitness-to-drive of older drivers can be best predicted. However, fitness-to-drive is affected by many factors, how to select the appropriate evaluation indicators is a problem that cannot be ignored. In addition, most current predictions of fitness-to-drive fall into two categories: fit and unfit for driving. In fact, older drivers will go through a transition process from fit to unfit for driving. In this period, older drivers can improve their fitness-to-drive through targeted cognitive and body training, this process is difficult to be defined as fit or unfit to drive. The traditional dichotomy of fitness-to-drive of older drivers is not suitable for the real situation, so it is necessary to add an intermediate classification in between, and driving intervention can then be targeted to this group of older drivers.

In this study, data from the functional ability test, the simulated driving test, and the on-road driving test of a group of older drivers are collected first, The prediction indicators suitable for three types of road test results are selected afterwards, and three machine learning techniques, i.e., decision tree, random forest and support vector machine, are applied to make classification prediction for the constructed indicator system.

2 Data Description

Research data came from 92 older drivers in Belgium, including 73 men and 19 women, they all aged over 70 but keep to drive. The evaluation took place over three days. All the participants took tests related to visual, physical and cognitive abilities at the hospital on the first day. On the second day, driving simulation tests were conducted in the Transportation Research Institute (IMOB) of Hasselt University under the guidance of researchers. On the third day, road tests were carried out and scored by professionals to determine whether they were fit to drive or not.

2.1 Functional Ability Tests

Functional tests were performed on the first day and were designed to assess visual, physical and cognitive function of each participant. These tests were chosen because they have previously been shown to correlate with driving ability.

In terms of Visual function, Snellen Chart and Pelli Robson were used to evaluate Visual acuity and Contrast sensitivity respectively [15].

In respect to motor function, the Timed get-up-and-go and the Functional Reach test are used to evaluate mobility [16] and physical balance [17] respectively.

Cognitive ability was evaluated from four aspects: overall cognitive ability, memory capability, attention and executive ability. Mini-mental state examination and clock drawing test are used to evaluate overall cognitive ability, rey complex figure test, eight word subtest, WAIS III - Digit Span are used to assess memory capacity. In terms of attention, trail making test (TMT, including A and B) and useful field of view (UFOV) are used to test visual processing speed, divided attention and selective attention ability, respectively.

Knowledge of road signs and the precursor maze test are used to assess executive ability, in which the knowledge of road signs is used to assess fitness-to-drive after a stroke, and the Maze test assesses visuospatial ability, planning and visual attention.

2.2 Simulated Driving Test

The simulated driving test was conducted in a stationary STISIM3 driving simulator at IMOB. The participants were operated under the guidance of professional researchers. They were required to adapt to the driving simulator by performing practice scenarios twice before the formal test. If there was any simulator disease, the test would be stopped.

The simulated driving includes six driving events that are difficult for older drivers. They are merging into traffic, maximum deceleration-stop sign, initial brake point-pedestrian crosswalk, turning left-gap acceptance, accidental crossing of pedestrians without precursor and accidental crossing of pedestrians with precursor. Driving simulation sets up three driving environments, namely rural road, urban road and motorway, with two different complexity levels, i.e. high traffic flow and low traffic flow.

After the simulated driving test, some important data were recorded or computed. To be specific, merging into traffic event took place on the motorway, and the distance (m) required from the on-ramp to the mainline was recorded. Maximum deceleration-stop sign event recorded the maximum deceleration (m/s^2) within 100 meters before the intersection with a stop sign. Initial brake point-pedestrian crosswalk event took place on rural road, and the distance from the position where participants starting braking to the zebra crossing was recorded. Turning left-gap acceptance event happened at urban road intersections, recorded the clearance time (s) when the participant decided to cross the traffic in front when turning left.

Accidental crossing of pedestrians without precursor event and accidental crossing of pedestrians with precursor event recorded detection time and reaction time. Detection time was time between hazard onset of movement pedestrian and throttle released by the participant for the first time, while reaction time was time between hazard onset of movement pedestrian and the first time participant hit the brakes. The difference was that the former happened in low traffic flow, the latter had a bus stop as a precursor and was in high traffic flow.

2.3 On-road Test

The on-road driving test was conducted in accordance with the official driver's license issuing procedures in Belgium, the TRIP score was made by professionals from the Center for Determination of Fitness to Drive and Car Adaptations (CARA) in the coach car. According to their driving performance and driving ability, all the participants were divided into three categories: fit to drive, undetermined and unfit to drive.

3 Methodology

3.1 Data pre-processing

After collecting the data from all the tests, the Z-score standardization method is applied to eliminate dimensional differences and outliers. The data processed by the Z-score method conform to standard normal distribution. It is suitable to the case that the threshold of the data is unknown and can help to remove those outliers.

The data processed by the Z-score method can take infinite value in theory, but in fact, the probability of the absolute value of the processed data greater than 3 is already very small. According to the standard normal distribution table, when the data conform to standard normal distribution, 99.74% of the data are within the range of ±3. Therefore, ±3 was selected as the threshold value of data after Z-score processing, and the standardized data with absolute value greater than 3 were treated as outliers and removed from the data set.

3.2 Indicator selection

In this study, the Pearson correlation coefficients between the functional test results and the road test results, as well as between the simulated driving results and the road test results, are calculated by using SPSS. The coefficients and their statistical significance are shown in Table 1 and Table 2.

Table 1 Correlation between functional test indicators and road test results

	Coefficient	Significance
Visual activity	0.290	0.005
Contrast Sensitivity	0.326 [1]	0.002*
Timed get-up-and-go	-0.278	0.007
Functional reach test	0.350 [1]	0.001*
MMSE	0.139	0.187
Clock drawing	0.112	0.286
Eight word subtest	0.196	0.061
Rey-copy	0.162	0.122
Rey-recall	0.200	0.056
Rey-delay	0.154	0.143
WAIS III – Digit Span-voor	0.102	0.335
WAIS III – Digit Span-acht	0.131	0.212
TMT-A	-0.339 [1]	0.001*
TMT-B	-0.263	0.011
UFOV1-Processing	-0.073	0.488
UFOV2-Divided attention	-0.347 [1]	0.001*
UFOV3-Selective	-0.389 [1]	0.000*
Maze Maximal number	0.203	0.052
Knowledge of road signs	0.322 [1]	0.002*

[1] | Pearson coefficient | > 0.3, * P < 0.005

According to the results in Table 1, we find that Contrast Sensitivity, Functional reach test, TMT-A, UFOV2-divided attention, UFOV3-selective and Knowledge of road signs, are statistically correlated with the results of the three on-road test categories (the p value<0.005).

Since UFOV2 and UFOV3 are highly correlated (Pearson coefficient=0.904, the p value=0.000), to avoid multicollinearity, UFOV3 with a larger correlation coefficient is retained. Thus, the remaining five indicators are selected as the predictive indicators representing these older drivers' functional ability.

Table 2 Correlation between simulated driving indicators and road test results

	Coefficient	Significance
Merging into traffic-distance	-0.179	0.087
Maximum deceleration–stop sign	0.130	0.217
Initial brake point–zebra crossing	0.045	0.673
Turning left–gap acceptance	0.089	0.397
Detection time to road hazards-no precursor	-0.311 [1]	0.003*
Detection time to road hazards-precursor	0.013	0.902
Reaction time to road hazards-no precursor	-0.263	0.011
Reaction time to road hazards-precursor	-0.096	0.362

[1] | Pearson coefficient | > 0.3, * $P < 0.005$

As can be seen from Table 2, Detection time to road hazards-no precursor is the only indicator from the simulated driving that has a strong correlation with the on-road test results. This indicator is therefore selected for further analysis.

The scatter diagram between the 6 indicators selected after screening and the on-road driving test results is shown in Figure 1.

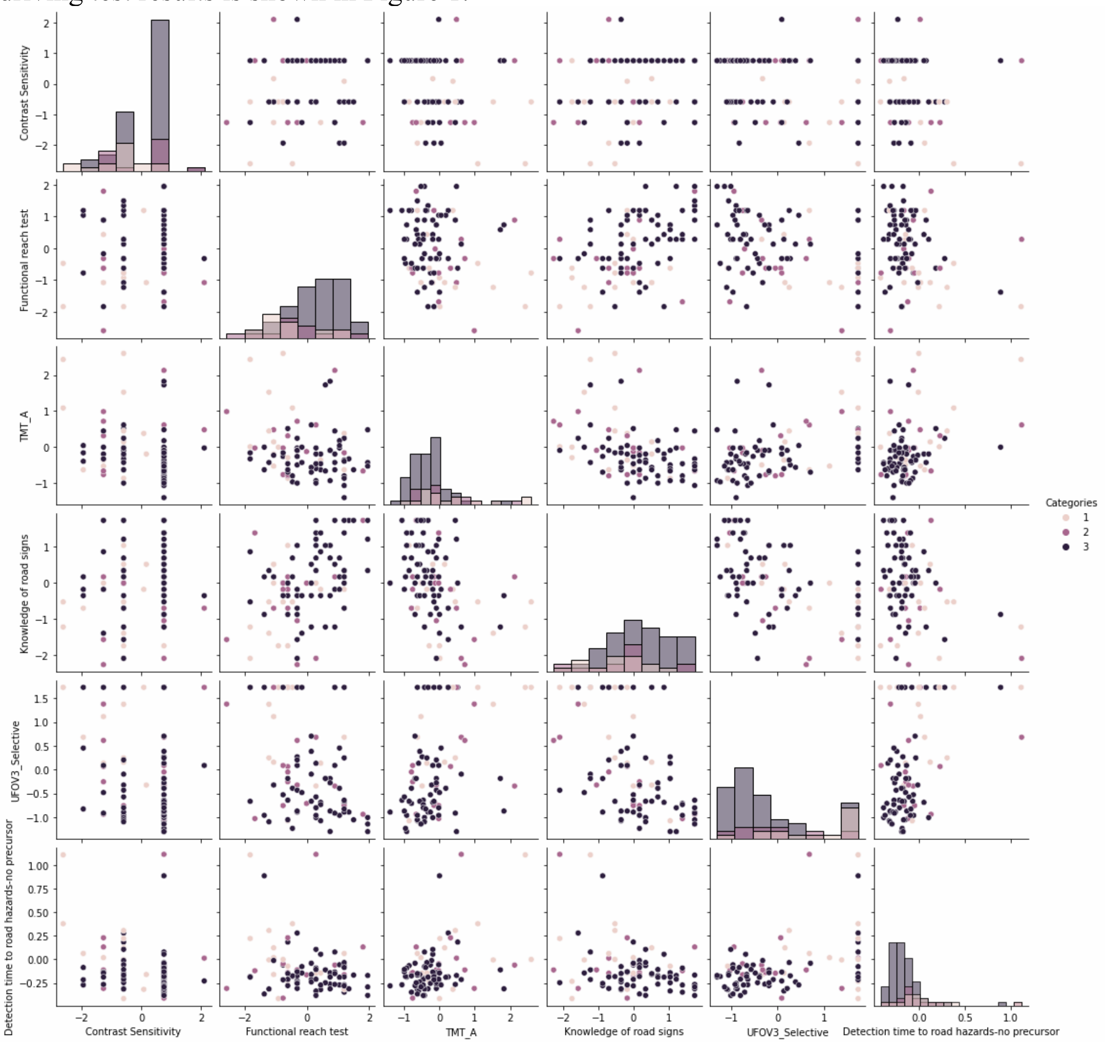

Figure 1 The data description scatter diagram of the selected indicators and road test results

As can be seen from Figure 1, there is no obvious linear relationship between the selected indicators and the three on-road test categories, and the outliers of the data after Z-score screening are controlled within a certain range with reasonable distribution.

3.3 Decision tree (DT)

Assessment of fitness-to-drive among older drivers is essentially a classification problem. In this paper, different from binary classification of fitness-to-drive in the previous studies, machine learning classification models will be used to classify the pre-processed data into three categories based on the selected six indicators.

Decision tree is established according to a given set of training data, making it possible to correctly classify data, and can choose an optimal solution under the condition of minimizing loss function. The essence of decision tree is to estimate conditional probability through training data set. It usually contains three steps: feature selection, decision tree generation and decision tree pruning. The computation is relatively small, but it is easy to cause over-fitting problems [18].

3.4 Random forest (RF)

Random forest is an integration method based on multiple decision trees, which can solve the over-fitting problem of single decision tree [19]. It sets the classification results of several decision trees, and takes the classification with the most votes as the final result. In other words, random forest combines several weak classifiers into a new strong classifier with better nonlinear fitting ability. However, random forest is not suitable for classifying data with small samples and low dimensions.

3.5 Support vector machine (SVM)

Support vector machine can classify the data of small sample well. Its core idea is to partition data sets by a hyperplane, which was originally used to solve dichotomous problems. In recent years, the proposal of one-against-all [20] and one-against-the-rest [21] makes it possible to solve multiple classification problems. But the process of applying the support vector machine to solve linear indivisible problem is complicated.

4 Application and Results

In this study, the 6 indicators introduced in Section 3.2 are taken as features, and the classification results from the on-road test are taken as labels. We randomly divide 70% of the data as training sets and 30% as test sets. Although dimensionality and outliers are removed from the original data, the problem of unbalanced data distribution is not solved. Among 92 participants, 66.3% were judged to be fit to drive, while only 16.3% were judged to be undetermined and 17.4% were unfit to drive. In order to improve the prediction accuracy, this paper chooses to carry out the technique of SMOTE [22] to balance sample on the training data, both the "unfit for driving" and "undetermined" categories were oversampled to the number of “fit to drive” in training data, which is consistent with the number of "suitable for driving" classification.

The default parameters of the DecisionTreeClassifier package in sklearn.tree are used for the DT model, and the default parameters of the RandomForestClassifier package in sklearn.ensemble are used for the RF model. In SVM classification, the nonlinear and dimensionality problems are considered when classification into three categories. The polynomial kernel function is used for dimensionality reduction, penalty parameter C=0.5 and balanced classification are selected. By running the three models for ten times each, we record the accuracy of the models, and compare the average accuracy afterwards.

Table 3 shows the results of ten classification predictions made by different models after learning the six indicators. The average accuracy of SVM is the highest, which is 75.00%. Followed by RF with the average accuracy of 71.79%. The prediction accuracy of DT is the lowest (67.50%) with the largest variance as well, which means that the prediction result from DT is not stable and will fluctuate up and down with the division of the training set and the test set. Thus, SVM is considered as a suitable approach for the assessment of older driver’s fitness-to-drive with multiple classification.

Table 3 The prediction results of three classification models

	DT (%)	RF (%)	SVM (%)
1	75.00	67.86	78.57
2	57.14	71.43	82.14
3	67.86	67.86	71.43
4	60.71	71.43	71.43
5	71.43	71.43	78.57
6	71.43	75.00	78.57

7	67.86	75.00	75.00
8	71.43	78.57	67.86
9	64.28	71.43	67.86
10	67.86	67.86	78.57
Mean (%)	67.50	71.79	75.00
Variance	29.64	12.60	25.49

The SVM model with the highest accuracy is selected for further analysis. Table 4 displays the confusion matrix of the predicted results. This prediction misclassifies two "fit to drive" drivers into "undetermined", two "undetermined" drivers into "fit to drive", and one "unfit to drive" driver into "fit to drive". The classification prediction of "unfit to drive" obtains an accuracy rate of 80%. In general, the SVM model achieves a reasonable result which largely avoids driving risk that may arise when the older drivers are misclassified.

Table 4 Confusion matrix of the second SVM prediction

	Unfit (predicted)	Undetermined (predicted)	Fit (predicted)
Unfit	4	0	1
Undetermined	0	0	2
Fit	0	2	19

5 Conclusion

In this study, by identifying the six indicators from the functional ability tests and the simulated driving test, which are contrast sensitivity, functional reach test, TMT-A, UFOV3-selective, knowledge of road signs and detection time to road hazards-no precursor, three machine learning methods, i.e., decision tree, random forest and support vector machine, were applied to conduct the fitness-to-drive assessment among a number of older drivers. Different from the traditional fitness-to-drive assessment, this study classified all the drivers into three categories and a transition classification “undetermined” was added between "fit to drive" and "unfit to drive". The result showed that the SVM model achieved the best performance, with the highest accuracy rate over 80%. Therefore, the assessment procedure proposed in this study can be considered as a valuable alternative way to the on-road test for the fitness-to-drive assessment for older drivers.

In the future, more aspects could be investigated. For instance, the role of each indicator from different tests in the final assessment can be explored, and the extension of the SVM model or some other more advanced machine learning methods can also be tried. More importantly, after the fitness-to-drive assessment, specific training plans should be proposed, especially for those “undetermined” older drivers, to improve their driving ability and to be fit to drive for a longer time.

Acknowledgements

This research was supported by the Open Project of Key Laboratory of Ministry of Public Security for Road Traffic Safety (Grant No. 2022ZDSYSKFKT01), and the Natural Science Foundation of Jiangsu Province of China (Grant No. BK20190371). The authors also want to thank colleagues from IMOB for their efforts on data collection.

References

[1] Piersma D, Fuermaier A, Waard D, et al. Prediction of fitness to drive in patients with Alzheimer's dementia[J]. Plos One, 2016, 11(2): 1-29.
[2] Meyers J E, Volbrecht M, Kaster-Bundgaard J. Driving Is More Than Pedal Pushing[J]. Applied Neuropsychology, 1999, 6(3):154-164.
[3] Lee H C, D Cameron, Lee A H. Assessing the driving performance of older adult drivers: on-road versus simulated driving[J]. Accident Analysis & Prevention, 2003, 35(5):797-803.
[4] AO Caffò, Tinella L, Lopez A, et al. The Drives for Driving Simulation: A Scientometric Analysis and a Selective Review of Reviews on Simulated Driving Research[J]. Frontiers in Psychology, 2020, 11:917.
[5] Haegerstrom-Portnoy G, Schneck M, Brabyn J. Seeing into old age: Vision function beyond acuity[J]. Optometry and Vision Science, 1999, 76(3): 141-158.
[6] Blane A. Through the looking glass: A review of the literature investigating the impact of glaucoma on crash risk, driving performance, and driver self-regulation in older drivers[J]. Journal of Glaucoma, 2014, 25(1): 113-121.
[7] Eby D, Molnar L, St. Louis R. Older adults, in: Perspectives and Strategies for Promoting Safe Transportation among Older Adults[M]. Elsevier, 2019, 73-100.
[8] Carr D B. Physician's guide to assessing and counseling older drivers[J]. Annals of emergency medicine, 2004, 43(6): 746-747.
[9] Korner-Bitensky N, Sofer S. The DriveABLE Competence Screen as a predictor of on-road driving in a clinical sample[J]. Australian Occupational Therapy Journal, 2010, 56(3): 200-205.
[10] Kay L G, Bundy A C, Clemson L M. Predicting Fitness to Drive in People With Cognitive Impairments by Using DriveSafe and DriveAware[J]. Archives of Physical Medicine & Rehabilitation, 2009, 90(9):1514-1522.
[11] Innes C, Jones R D, Dalrymple-Alford J C, et al. Sensory-motor and cognitive tests predict driving ability of persons with brain disorders[J]. Journal of the Neurological Sciences, 2007, 260(1-2):188-198.
[12] McCarthy D P, Mann W C. Sensitivity and specificity of the assessment of driving-related skills older driver screening tool[J]. Topics in Geriatric Rehabilitation, 2006, 22(2): 139-152.
[13] Shen Y J, Zahoor O, Tan X, et al. Assessing fitness-to-drive among older drivers: A comparative analysis of potential alternatives to on-road driving test[J]. International Journal of Environmental Research and Public Health, 2020, 17(23): 1-18.
[14] Shen Y J, Tan X, Brijs T. Assessment of fitness-to-drive for elder drivers based on functional tests and simulated driving test[J]. Journal of Southeast University: Natural Science edition, 2021,1:171-177.
[15] Pelli D G, Robson J G, Wilkins A J. The design of a new letter chart for measuring contrast sensitivity[J]. Clinical Vision Sciences, 1988, 2(3):187--199.
[16] Podsiadlo D, Richardson S. The Timed "Up & Go": A Test of Basic Functional Mobility for Frail Elderly Persons[J]. Journal of the American Geriatrics Society, 1991, 39(2):142-148.
[17] Duncan P W, Weiner D K, Chandler J, et al. Functional Reach: A New Clinical Measure of Balance[J]. Journal of Gerontology, 1990, 45(6): M192–M197.
[18] Bansal M, Goyal A. and Choudhary A. A comparative analysis of k-nearest

neighbour, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning[J]. Decision Analytics Journal, 2022, [online] p.100071.
[19] Rakhra M, Soniya P, Tanwar D, et al. Crop price prediction using random forest and decision tree regression: -a review[J]. Materials Today: Proceedings, 2021.
[20] Hs U C, Lin C J. A comparison of methods for multiclass support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2): 415-425.
[21] Weston C J. Multi-class support vector machines[C]. Springer London. Springer London, 2005.
[22] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: Synthetic Minority Over-sampling Technique[J]. Journal of Artificial Intelligence Research, 2002, 16(1):321-357.