



Jibin Yin<sup>1</sup>, Pengfei Zhao<sup>1</sup>, Yi Zhang<sup>2,\*</sup>, Yi Han<sup>3</sup> and Shuoyu Wang<sup>3</sup>

- <sup>1</sup> Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; yinjibin@kust.edu.cn (J.Y.); zhaopengfei@stu.kust.edu.cn (P.Z.)
- <sup>2</sup> Naval Training Base of Health Service, Naval Medical University, Shanghai 200433, China
- <sup>3</sup> School of System Engineering, Kochi University of Technology, Kami City 780-8520, Japan;
- 258012g@gs.kochi-tech.ac.jp (Y.H.); wang.shuoyu@kochi-tech.ac.jp (S.W.)
- \* Correspondence: 18602160005@126.com

**Abstract**: The demand for large-scale analysis and research of data on trauma from modern warfare is increasing day by day, but the amount of existing data is not sufficient to meet such demand. In this study, an integrated modeling approach incorporating a war trauma severity scoring algorithm (WTSS) and deep neural networks (DNN) is proposed. First, the proposed WTSS, which uses multiple non-linear regression based on the characteristics of war trauma data and the medical evaluation by an expert panel, performed a standardized assessment of an injury and predicts its trauma consequences. Second, to generate virtual injury, based on the probability of occurrence, the injured parts, injury types, and complications were randomly sampled and combined, and then WTSS was used to assess the consequences of the virtual injury. Third, to evaluate the accuracy of the predicted injury consequences, we built a DNN classifier and then trained it with the generated data and tested it with real data. Finally, we used the Delphi method to filter out unreasonable injuries and improve data rationality. The experimental results verified that the proposed approach surpassed the traditional artificial generation methods, achieved a prediction accuracy of 84.43%, and realized large-scale and credible war trauma data augmentation.

Keywords: artificial intelligence; data augmentation; war trauma severity score; deep neural network

## 1. Introduction

War trauma data are the core elements of wargaming, military medical service training, and medical decision-making [1]. With the continuous development of modern warfare, the analysis and research of physical war trauma data have become more and more important. However, the amount of existing data is not sufficient to support large-scale analysis and evaluation, and the confidential nature of war trauma data makes them hard to collect and obtain from public channels. Therefore, efficient and credible data augmentation of war trauma data has become a research work with great practical significance. To the best of our knowledge, research on this topic has been limited. In the currently used method, the additional physical trauma data are still artificially generated by well-trained experts or doctors based on their professional knowledge and experience. However, this method is not only inefficient, time-consuming, and labor cost-intensive, but also inherently biased due to its dependence on personal subjective cognition, which is difficult to overcome. In addition, different experts have no unified standard for assessing injury consequences. Furthermore, the amount of artificially generated war trauma data is too small to meet the actual needs. Therefore, we developed a standardized evaluation algorithm to improve the quality of assessment of injury consequences and find an automatic, efficient, and credible approach for small-sample augmentation of war trauma data.

More than half a century since the concept of artificial intelligence (AI) was first formally proposed at the Dartmouth Conference [2], the AI technology has empowered



Citation: Yin, J.; Zhao, P.; Zhang, Y.; Han, Y.; Wang, S. A Data Augmentation Method for War Trauma Using the War Trauma Severity Score and Deep Neural Networks. *Electronics* **2021**, *10*, 2657. https://doi.org/10.3390/ electronics10212657

Academic Editor: Enzo Pasquale Scilingo

Received: 16 September 2021 Accepted: 26 October 2021 Published: 29 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). amazing developments in many fields. Meanwhile, the external environment and challenges faced by the development of AI have also undergone profound changes [3]. These changes are especially prominent in certain fields, such as big data, virtual reality, supercomputing, and mobile payment. Therefore, under the trend that the overall environment is getting closer to big data, deep learning (DL), which is based on machine learning, has become the core element of the application of AI [4] and has led to satisfactory application results in many fields, such as cloud computing [5], image identification [6], sports training [7], and AlphaGo [8]. Recently, AI technologies such as DL started to be gradually applied in the field of medical research, including in promoting disease management [9], computer-aided diagnosis [10], biomedical information processing [11], medical image recognition [12], and disease prediction [13]. Especially in disease prediction, AI has been recognized as one of the key elements of an accurate and robust prediction system [14]. For example, deep neural networks (DNNS), which are AI tools, are now used to assist physicians and for automatic diagnosis. Specific application cases include early detection of cardiovascular disease [15], cancer diagnosis [16], survival prediction [17], and injury severity assessment [18].

Compared with machine-learning methodologies and shallow neural networks, DL, which is now the core of the AI method, overcomes the research drawbacks of limited samples and low generalizability by training large-scale annotated sample data to automatically extract complex sample features and fully optimize the model parameters layer by layer. Thus, DL can carry out a more essential characterization of the data and demonstrates a superior feature-learning ability [19]. In other words, with the existing technology level, the larger the scale and the higher the quality of the annotated data are, the better the performance of the model will be. Therefore, DL can effectively solve many complex problems in the medical field [20,21]. In the prediction and diagnosis of some diseases, the accuracy and efficiency of predictive DL models have surpassed those of professional doctors and experts [22] and have thus made outstanding contributions to the development of the medical field.

### 2. Related Work

Currently, there are two main methods of data augmentation: oversampling and generative adversarial network (GAN). The principle of oversampling is as follows: if the samples of different classes are imbalanced, the training data can be expanded by copying the training samples of the minority class or adding noises to create new ones [23]. To solve the imbalanced dataset learning problem, in 2002, Bowyer et al. [24] created a synthetic minority oversampling technique (SMOTE), which generated synthetic minority class samples. In 2005, Han et al. [25] proposed a borderline SMOTE algorithm, which considered the minority instances near the borderline and the neighboring instances. The following year, David et al. [26] proposed a cluster SMOTE; Bai et al. [27] proposed an adaptive synthetic sampling approach (ADASYN) for imbalanced learning in 2008; Barua et al. [28] suggested a MWMOTE in 2014; Douzas et al. [29] proposed a SOMO method in 2017. Most of these methods focused on imbalanced learning by adding oversampling examples to the imbalanced datasets. However, physical war trauma data are not imbalanced but insufficient in every class. Therefore, the abovementioned oversampling techniques are not suitable for the augmentation of physical war trauma data.

A GAN is a data augmentation model based on DL, which can be used to learn the potential distribution of complex data, generate large-scale and high-quality synthetic samples, effectively solve the problem of insufficient data due to factors such as difficulty and cost of sample acquisition [30]. Thus, the GAN has become one of the most promising data augmentation approaches in recent years. A GAN is intrinsically a generation model [31] that does not depend on a priori hypotheses but on the internal confrontation between the data and the model itself to achieve unsupervised learning. To solve the inadequate problem of real data, a GAN can generate synthetic samples of the existing data with the same distribution [32]. A GAN's structure consists of two feedforward neural networks:

a generator G and a discriminator D. In the learning process, G continuously generates new synthetic samples while D discriminates between the synthetic samples and the real samples as accurately as possible, then gives feedback. In this way, the GAN has created a game similar to "counterfeit currency identification" in which both sides of the game continue to improve their abilities through confrontation.

However, the samples processed by a GAN are mainly two-dimensional data such as pictures and voice signals. A GAN generates virtual images by rotating, scaling, cropping, and changing the brightness, contrast, hue, saturation and adding random noise to image data. However, the GAN is not a good choice for augmenting physical war trauma data.

In the medical field, the application of medical scoring is increasingly maturing, especially in medical treatment, early diagnosis, trauma assessment, and other aspects to the point that it now plays an important auxiliary role. For example, Gabriele Canzi et al. introduced the comprehensive facial injury (CFI) score for comprehensively evaluating severity of facial injuries [33]. Hasanka Ratnayake et al. used a laboratory-derived early warning score to predict in-hospital mortality and admission to the intensive care unit (ICU) [34]. Konlawij Trongtrakul et al. created the acute kidney injury (AKI) risk prediction score for early prediction of the condition among critically ill surgical patients [35].

The trauma score is a common type of medical score that predicts severity of an injury. It uses scientific scoring to quantitatively or semi-quantitatively assess injury severity and its consequences to the injured [36]. The scoring standard was developed by a panel of experts in the field who will continue to improve and optimize it based on feedback from the application of the trauma score as well as from related research progress. Recently, several improved injury severity score (ISS) methods have been proposed. Cristiane et al. created a novel trauma and injury severity score (TRISS) for survival prediction [37]; Yang et al. used a revised injury severity score (RISS) to evaluate the severity of injuries of patients hospitalized due to an accident [38]; Shi et al. developed a weighted injury severity score (WISS) to improve adult trauma mortality prediction [39]. For example, RISS divides the human body into six public parts: the head, the face, the chest, the abdomen, the limbs, and the body surface. Then, it squares the standard ISS for each of the most serious injuries of the three most serious body parts of the patient and puts them together. As for the second most serious injuries, only their ISS values are put together. If there are more than four injured parts, the standard injury severity score of the most serious injuries of the fourth part is added. The RISS equation is as follows:

$$RISS = (A_1^2 + A_2) + (B_1^2 + B_2) + (C_1^2 + C_2) + D$$
(1)

where  $A_1$ ,  $B_1$ , and  $C_1$  mean the most serious injuries of the three most serious body parts;  $A_2$ ,  $B_2$ , and  $C_2$  mean the second most serious injuries of the three most serious body parts; D means the standard injury severity score of the most serious injuries of the fourth part.

Taken together, various novel scientific scoring methods have gradually become doctors' helping hands in evaluating patients' injuries. Medical scoring belongs to the category of predictive science. Because different scoring mechanisms have different limitations, it is impossible to achieve 100% accuracy in prediction. However, with the continuous advancements in medicine and with the revision, expansion, and improvement of the scoring mechanisms by researchers in the related domains, medical scoring approaches are expected to become more scientific, practical, and in line with objective reality [40].

On the other hand, the DL technology combined with knowledge from different disciplines for interdisciplinary field research is an emerging trend. For example, Yang et al. enhanced PIR-based multiperson localization by combining DL with the domain knowledge [41], and Ding et al. combined the domain knowledge and DL for domain adaptation in machine translation [42]. Therefore, combining DL with the domain knowledge of medical experts according to the characteristics of war trauma data is key to the successful application of DL to the augmentation of war trauma data.

Based on the above research, to solve the data augmentation problems with smallsample war trauma data by studying the GAN's idea and the medical trauma scoring method, this article proposes an approach that combines a WTSS with a DNN [43]. The WTSS–DNN integrated model simulates the generative model in thought, including sample generation and discrimination. The injuries are generated through random sampling and evaluated with WTSS, and then marked with an injury consequence label; this is the sample generation link. The assessment of the prediction accuracy of the DNN classifier is combined with the discrimination of unreasonable injuries by the expert panel; this is the discrimination link. After the accuracy and plausibility of the synthetic samples have been judged, the expert panel provides feedback, based on which, on the one hand, the characteristics of the synthetic samples are further investigated while the necessary optimization and adjustments to the WTSS algorithm are made; and on the other hand, the unreasonable synthetic samples are filtered out to improve data rationality. Eventually, the accuracy and plausibility of the augmented data are expected to stabilize and be optimized to generate credible samples.

This data augmentation approach is the first attempt to combine war trauma assessment in the medical field with DL in the AI field. The WTSS–DNN integrated model can automatically generate large-scale and credible virtual war trauma data, making it possible to carry out related data-based military research, which has great practical significance. In addition, this approach not only helps to solve the war trauma data augmentation problem, but the WTSS algorithm we have proposed also provides a practical auxiliary tool for quickly evaluating soldiers' injuries and formulating treatment strategies.

### 3. Materials and Methods

In this section, we first explain the overall process of the research, then introduce the WTSS algorithm in detail. Next, we introduce the structure of our DNN classifier, and then determine the multiclassification metrics used in the algorithm to evaluate the performance of the classifier. Finally, the method of judging the plausibility of the generated synthetic samples is introduced.

#### 3.1. Workflow of the Study

To solve the data augmentation problem and the supervised learning problem, an integrated modeling approach that incorporates the war trauma severity scoring algorithm (WTSS) and a DNN model was proposed. This approach's workflow is summarized as follows (Figure 1).

- 1. Based on the known probability distributions, the injured parts, injury types, and complications were randomly sampled and then combined to form a complete war trauma injury condition. Next, we used the WTSS algorithm to calculate the severity score and evaluate the consequences, after which the injury consequence label was marked.
- 2. After the data preprocessing, to test the accuracy of the injury consequence prediction, we trained a DNN classifier with the generated data and tested it with real data.
- 3. Through the Delphi method, the expert panel reached a consensus on unreasonable multiple injuries based on the domain knowledge [44] and then filtered out the unreasonable synthetic samples after the data generation.
- 4. After the predicted accuracy was evaluated and the unreasonable synthetic samples were filtered out, credible virtual war trauma data were finally output.

#### 3.2. Random Injury Generation

In the injury generation process, we first randomly sampled the injured part according to the probability of occurrence; then, we randomly selected the possible injury types according to the injured part; finally, we randomly sampled whether it is accompanied by complications; if there were complications, we randomly selected the possible complications.



Figure 1. Workflow of the WTSS-DNN integrated approach.

## 3.3. WTSS Algorithm

After injuries were randomly generated, the focus of the research was on how to conduct standardized and accurate injury assessments. To solve this problem, we conducted multiple rounds of discussions and communication with the expert panel and finally decided to carry out a standardized quantitative assessment of various injuries by proposing a war trauma severity scoring algorithm.

Via in-depth summary of the various existing trauma scoring algorithms and based on the idea of multiple nonlinear regression and the key factors that affect severity of an injury (injured part, injury type, complications, and whether there are multiple injuries), after several rounds of testing and optimization, the equation for WTSS was finally determined as follows:

$$F(P, X, C) = a + \sum_{i=0}^{6} P_i X_i + C_i$$
(2)

where *F* represents the severity score;  $P_i$  represents the weight coefficient of injury severity for each of the seven body parts;  $X_i$  shows whether the corresponding body part was injured (if not injured, the corresponding  $X_i$  value equals 0; otherwise, it equals the injury severity standard score for the corresponding body part);  $C_i$  shows whether the injury was accompanied by complications (if there were no complications,  $C_i$  equals 0; otherwise, it equals the corresponding severity score); the bias *a* is the correction value for multiple injuries (if there were multiple injuries, *a* equals –20; otherwise, it equals 0).

Next, we calculated *F* according to the predictive factors  $P_i$ ,  $X_i$ ,  $C_i$ , and *a*, then selected the corresponding score interval according to the magnitude of *F*. Finally, we labeled the synthetic samples with the consequences of the injury. The pseudocode of WTSS is provided in Algorithm 1.

```
Algorithm 1. War trauma severity score (WTSS).
Input: Weight coefficient of injury parts: P_i = \{P_0, P_1, ..., P_6\}.
       Injury type score: X_i = \{X_0, X_1, ..., X_6\}.
       Complication score: C_i = \{C_0, C_1, ..., C_6\}.
       Correction value for multiple injuries: a = -20.
Output: Severity score: F(P, X, C).
1: n = 0
2: for i = 0 to 6 do
     if P_i \neq 0 and X_i \neq 0 then
3:
4:
        F(P, X, C) += P_i * X_i
5:
        n += 1
6:
     end if
7:
     if C_i \neq 0 then
8:
       F(P, X, C) += C_i
9:
     end if
10: end for
11: if n > 1 then
12: F(P, X, C) += a
13: end if
14: return F(P, X, C)
```

The WTSS algorithm is a nonlinear model which ignores complicated details of the injury and uses a good correlation between the injuries' consequences and the severity of the injured parts and the injury types [45]. The weight coefficients of injuries in different body parts are shown in Table 1, and the example of the standard severity score for injury types and complications are shown in Figures 2 and 3. The score intervals for the injury consequences are listed in Table 2.

Table 1. Weight coefficients of each body part.

Body Part	Weight Coefficient
Head	8
Face	8
Neck	8
Chest / back	7
Abdomen	6.5
Pelvis / hip	6.5
Limbs	5

In a situation wherein different injury types or complications have the same standard injury severity score in a certain injured part, we coded them to distinguish. Taking the abdomen as an example, the coding method is shown in Figure 4.

As an independent scoring algorithm to determine severity of war trauma, WTSS does not perform an extremely accurate diagnosis of a specific injury. Instead, it performs standardized assessment and prediction of the most probable consequences of injuries from an objective perspective to ensure accuracy of the injury consequence assessment. Additionally, WTSS is not only the core of our WTSS–DNN integrated model that contributes to large-scale analysis and evaluation of war trauma data, but it also helps to quickly evaluate and diagnose soldiers' injuries on the battlefield and determine the treatment strategy. Furthermore, in complex battlefield environments, the soldier's age, physical constitution, and other factors may cause different consequences of the same trauma. Consequently, WTSS only objectively assesses the injury without considering the age and other physiological indicators to meet the requirements of the ideal scoring method that is "easy to implement, objective, and accurate" [38].

Injury part	Injury t <b>ype</b> (Score)			
Head	Perforating injury I (8) Perforating injury II (11)	Penetrating injury I (3.5) Fracture I (3) Penetrating injury II (10) Fracture II (3)		
Face	Perforating injury I (8) Perforating injury II (7)	Penetrating injury I (6) Fracture I (3) Penetrating injury II (6) Fracture II (3)		
Neck	Penetrating injury II (7)	Soft tissue injury I (5) Soft tissue injury II (4)		
Chest / Back	Perforating injury I (7) Perforating injury II (8.5)	Penetrating injury I (3) Penetrating injury II (8) Severe fracture I (7.5)		
Abdomen	Perforating injury I (7) Perforating injury II (7)	Penetrating injury I (8) Penetrating injury II (7.5) Tangential injury I (3)		
Pelvis / Hip	Perforating injury I (9) Perforating injury II (8)	Penetrating injury I (8) Penetrating injury II (7.5)		
Limbs	Perforating injury I (7) Perforating injury II (5)	Penetrating injury I (3) Limb mutilation I (6.5) Tangential injury I (3) Penetrating injury II (3) Limb mutilation II (6) Tangential injury II (2)		

**Figure 2.** Standard severity score for different injury types. In this Figure, I indicate that the injury is a blast injury, II indicates that the injury is a gunshot wound.

Injury part	Complication (Score)				
Head	Major bleeding c Major bleeding c	caused by a penetrating injury (20) caused by a severe fracture (30)	Brain hernia (10) Increased intracranial pr	essure (7)	
Face	Shock (15)	Blindness (15)			
Neck	Dyspnea (20)	Hemorrhagic shock (25)			
Chest / Back	Major bleeding caused by a penetrating injury (10) Major bleeding caused by a perforating injury (15) Hemopneumothorax (30)		Hemorrhagic shock (25) Hemothorax shock (25) Cardiac rupture (40)	Pneumothorax (5) Hemothorax (20)	
Abdomen	Major bleeding caused by a penetrating injury (15) Hemorrhagic shock (25)		Shock (10) Peritonitis (3)		
Pelvis / Hip	Shock (15)	Hemorrhagic shock (20)			
Limbs	Shock (15)	Hemorrhagic shock (25)	Major bleeding (15)		

Figure 3. Standard severity score for different complications.

 Table 2. Description of the score intervals.

Score Interval	Consequence	Label
0–25	Minor injury	1
26–50	Moderate injury	2
51–75	Serious injury	3
75+	Critical injury and death	4



Figure 4. Example of injury coding in the abdominal area.

## 3.4. Deep Neural Network

Because the WTSS algorithm is a complicated nonlinear model, this article used a DNN as a classifier model to test the accuracy of injury consequences. The DNN classifier consists of an input layer, an output layer, and several hidden layers. It uses multilayer nonlinear information processing, which can be widely and flexibly used to solve problems such as classification, regression, dimensionality reduction, feature extraction, and clustering. First, we built a suitable DNN classifier network structure according to the actual needs, and the network structure was determined to be 22-16-16-16-4 after the experiment. Next, to test whether such a classifier has excellent generalization ability, we trained it with synthetic samples and tested it with real samples. To verify its performance, we used four multiclassification metrics based on a confusion matrix: accuracy, precision, recall, and the F<sub>1</sub> score [46]. Among these metrics, the F<sub>1</sub> score is the harmonic average of precision and recall. Finally, we adjusted and optimized the hyperparameters and then determined the best learning rate and the training sample size. The confusion matrix is shown in Figure 5.



Figure 5. Graph of the multiclassification confusion matrix.

In Figure 5, *L* represents the class number,  $n_{ii}$  and  $n_{ij}$ —the number of class  $C_i$  samples correctly predicted as class  $C_i$  and incorrectly predicted as class  $C_i$ , respectively;  $R_i$  and  $P_i$  indicate the recall and the precision of class  $C_i$ , defined in Equations (3) and (4), and the accuracy and the  $F_1$  score are defined in Equations (5) and (6).

$$P_i = \frac{n_{ii}}{\sum\limits_{j=1}^{L} n_{ji}}$$
(3)

$$R_i = \frac{n_{ii}}{\sum\limits_{j=1}^{L} n_{ij}} \tag{4}$$

$$Accuracy = \frac{\sum_{i=1}^{L} n_{ii}}{\sum_{i=1,i=1}^{L} n_{ij}}$$
(5)

$$F_{1} \ score = 2 \frac{\sum_{i=1}^{L} R_{i} \sum_{i=1}^{L} P_{i}}{\sum_{i=1}^{L} R_{i} + \sum_{i=1}^{L} P_{i}}$$
(6)

### 3.5. Discrimination of Unreasonable Injuries Based on the Delphi Method

After data generation, to improve the data plausibility of the synthetic samples, the expert panel reached a consensus on multiple unreasonable injuries based on the domain knowledge and provided feedback. Based on this feedback, we analyzed the law of unreasonable injury combinations and filtered out the unreasonable synthetic samples to improve data plausibility. Finally, we outputted the credible synthetic samples.

### 4. Empirical Analysis

Due to the high confidentiality and difficulty of access to war trauma data, it is gradually attracting greater attention from the army, military academies, and related hospitals. To eliminate obstacles to related research, an efficient and credible data augmentation approach is urgently needed in order to support large-scale war trauma data research and war game deduction. Our proposed integrated model provides a new and feasible way to meet the real need for large-scale and automated generation of credible war trauma data.

#### 4.1. Data Collection

In this study, we collected and organized two types of real war trauma data at a certain scale: data on gunshot wounds and blast injuries. We selected 338 cases (minor injury, 114 cases; moderate injury, 82 cases; serious injury, 74 cases; and critical injury and death, 68 cases) complete with the available data to form the test set. After the preprocessing operations such as one-hot encoding, data standardization, and feature reduction, our war trauma data had a total of 22 features.

#### 4.2. Results Analysis

We implemented our proposed WTSS–DNN integrated model in Python 3.7.7 and conducted experiments on a personal computer with a Windows 64-bit operating system. After a series of tests on the DNN, the optimal values of all the hyperparameters were determined. The classifier's input dimension was 22, equal to the feature dimension of the war trauma samples. The number of hidden layers of the classifier was set at 4, with each using ReLUs as the activation function. The softmax function was used as the output layer, and categorical cross-entropy was used as the loss function. We used TensorFlow 2.0.0 and GPU to train our DNN classifier; the epoch was set at 1000 and the batch size was set at 256. We chose Adam as our optimization algorithm as it performed best compared to SGD and RMSProp3 [47].

After determining the best network structure of the DNN classifier (22-16-16-16-16-4), we conducted contrast experiments at different learning rates [48]. Specifically, we kept the network structure and other hyperparameters unchanged, then set the values of the learning rate to be 0.05, 0.02, 0.01, 0.005, 0.002, 0.001, 0.0005, and 0.0001, respectively. Table 3 shows accuracy, precision, recall, and the F<sub>1</sub> score at different learning rates on the

same training set with a sample size of 10,000. The results show that the 0.001 learning rate led to the best overall model performance and thus was selected and used.

LR	Accuracy	Precision	Recall	F <sub>1</sub> Score
0.05	48.82	48.89	51.11	49.98
0.02	49.41	49.47	51.85	50.63
0.01	59.99	67.35	66.59	66.97
0.005	65.43	72.51	71.15	71.82
0.002	73.19	78.46	80.37	79.40
0.001	81.57	88.08	87.70	87.89
0.0005	78.33	86.33	85.74	86.03
0.0001	73.37	79.57	80.85	80.20

Table 3. Comparison of the multiclassification metrics at different learning rates.

Next, we explored the best training sample size (*n*). On the one hand, low numbers of training samples cannot fully teach sample features and meet the requirements of model accuracy; and on the other hand, too high numbers of training samples can increase the calculation costs and time costs and are not conducive to optimizing the hyperparameters. Therefore, we sought to determine the best training sample size in the range of 1000–20,000 through the trial and error method [49]. In the search process, to avoid the impact of class imbalance on the experimental results, synthetic samples of the four classes were extracted at the same proportion to form a training set for the experiment and test. The overall performance results of the multiclassification metrics at different training sample sizes are shown in Table 4.

п	Accuracy	Precision	Recall	F <sub>1</sub> Score
1000	73.17	78.51	79.00	78.75
2000	75.15	81.67	80.96	81.31
4000	78.99	85.14	84.70	84.91
8000	80.88	87.97	86.81	87.39
12,000	84.33	90.07	88.44	89.25
16,000	82.36	89.07	89.10	89.08
20,000	80.18	87.21	88.33	87.77

Table 4. Comparison of the multiclassification metrics at different sample sizes.

The experimental results showed that the small-scale training set did not meet the requirements for model accuracy. As the training sample size continued to increase, the predicted accuracy gradually increased. When the training sample size was 8000, the accuracy reached 80.88%; and when the training scale increased to 12,000, the accuracy increased to 84.33%. However, model performance became deteriorated when the training scale was greater than 12,000, which indicates that blindly increasing the training scale could not guarantee a consistently higher classification accuracy. Besides, when the training scale was increased, as the harmonic average of precision and recall, the trend of the  $F_1$  score was basically consistent with that of accuracy. Therefore, we supposed that selecting a training sample size of 12,000 can achieve the best compromise between the training cost and the classification performance.

Finally, our DNN classifier achieved the best overall performance with 84.33% accuracy, 90.07% precision, 88.44% recall, and an 89.25% F<sub>1</sub> score.

## 4.3. Evaluation of WTSS Combined with a DNN

In this section, we first explored the accuracy of injury assessment of different classifier models. Subsequently, to evaluate the respective contributions of the WTSS algorithm and the DNN classifier in the WTSS–DNN data augmentation method, we set up an ablation

experiment. Finally, we provided the prediction results of the DNN for real data through the confusion matrix.

First, we compared our DNN model with three classic machine-learning classifiers: random forest (RF) [50], XGBoost [51], and naïve Bayes (NB) [52].

The RF, XGBoost, and NB models and our DNN model were trained with the same training set and then tested with the same real samples. As shown in Figure 6, our DNN classifier performed better than the three classic machine-learning models. The NB model showed the weakest performance in comparison with the other classifier models because when the number of features is large or when the correlation between the features is high, the NB classification effect is poor. These results indicate that classic machine-learning models cannot be effectively trained when there are few samples and verified that a DNN classifier trained with a large amount of data has better classification performance.



Figure 6. Performance of the different classification strategies.

Next, to evaluate the respective contributions of the WTSS algorithm and the DNN classifier in the WTSS–DNN integrated model, we set up an ablation experiment. Specifically, we combined different injury assessment methods with different classifier models to observe performance of various combinations. Injury assessment methods include the WTSS algorithm and the manual assessment method (MA); classification models include DNN, RF, XGBoost, and NB. The results of the ablation experiment are shown in Table 5.

Table 5. Ablation experiment of different injury assessment methods and classifier models.

Assessment Method	Classifier Model	Accuracy	Precision	Recall	F <sub>1</sub> Score
	RF	69.39	70.21	70.47	70.34
244	XGBoost	68.86	69.12	70.07	69.59
МА	NB	53.86	55.20	53.41	54.29
	DNN	71.24	72.66	70.81	71.72
	RF	81.46	85.06	83.01	84.02
MEGO	XGBoost	82.08	82.00	85.07	83.51
W155	NB	64.08	68.60	63.56	65.98
	DNN	84.33	90.07	88.44	89.25

From the results of the ablation experiment, we can see that the WTSS algorithm is better than the traditional manual evaluation method, the prediction performance of the DNN classifier is better than that of the machine-learning model, and the combination of WTSS and the DNN performs best. Therefore, the combination of WTSS and the DNN can effectively solve the data augmentation problem of war trauma data and shows superiority compared with artificial generation methods.

Finally, we provided the prediction results of the DNN for real data through the confusion matrix.

From Table 6, we can see that the prediction accuracy for minor injuries and moderate injuries is very high, but the prediction accuracy for critical injuries is only about 60%, which is caused by the complexity of critical injuries.

Predicted	Minor Injury	Moderate Injury	Serious Injury	Critical Injury and Death	Total
Keal					
Minor injury	114	0	0	0	114
Moderate injury	3	73	6	0	82
Serious injury	0	7	60	7	74
Critical injury and death	0	10	20	38	68
Total	117	90	86	45	338

Table 6. Confusion matrix of injury consequence identification.

# 4.4. Data Filtering

The Delphi method, also known as the "expert investigation method", was invented in 1946 by RAND Corporation in the United States. The Delphi method is based on the key assumption that predictions from groups are usually more accurate than predictions from individuals. The goal of this method is to use a structured iterative approach to obtain consensual opinions from an expert panel [44].

For the multiple injuries data generated, some injury combinations are unreasonable they are almost impossible to appear in a real war. To improve plausibility and usability of the synthetic samples in our experiment, we decided to use the Delphi method to evaluate unreasonable multiple injuries and filter them out. After several rounds of identification and discussions, the expert panel reached a consensus on the unreasonable multiple injuries based on the domain knowledge. We analyzed the experts' feedback and then filtered out the unreasonable synthetic samples to improve data plausibility to output credible samples. Next, to verify whether the data plausibility improved or not, we randomly selected 300 original multiple-injury synthetic samples and 300 filtered ones, put them into three groups, and conducted contrast experiments. Then, we counted the number of reasonable samples before and after filtering. The experimental results are shown in Figure 7.



Figure 7. The numbers of reasonable samples in the synthetic samples.

The experimental results showed that data plausibility of the synthetic samples filtered out was significantly improved in comparison with that of the original ones and came close to 100%.

# 5. Discussion

For the WTSS–DNN integrated model, plausibility and effectiveness of the WTSS algorithm play a crucial role in the performance of WTSS–DNN. Therefore, we evaluated plausibility and effectiveness of the WTSS algorithm through the two methods described below. First, the expert panel intervention and assistance. The parameter setting and the scoring standard of the algorithm were determined after multiple rounds of discussions and evaluations with the expert panel, which is highly reasonable and professional. Second, we tested plausibility and effectiveness of the algorithm through ablation experiments. In the ablation experiments, on the one hand, we used the DNN classifier to verify accuracy and plausibility of the algorithm in injury assessment. The experimental results show that the prediction accuracy rate reached 84.43%, which is a satisfactory result. On the other hand, we compared the WTSS algorithm with the traditional manual assessment. Therefore, compared with the artificially generated methods, the performance of the proposed WTSS algorithm combined with a DNN in war trauma data augmentation is superior, can ensure high data quality, and automatically generates large-scale war trauma data on demand.

However, the experiment also showed that the prediction accuracy of the severity of multiple injuries was lower than that for a single injury due to the complexity of multiple injuries. Furthermore, after determining the WTSS standards, the proposed approach no longer relies on additional professional knowledge due to the characteristics of DL. Thus, for nonprofessionals, the proposed approach has a low barrier to successful application. Although we were able to generate credible virtual trauma data only for blast injuries and gunshot wounds in this study, with the continuous real data collection, the types of war trauma we can generate will become more abundant. Finally, the combination of DL with medical scoring algorithms can be used for other types of injury data augmentation, such as for surgical injuries and emergency injuries.

## 6. Conclusions

In this article, the WTSS algorithm combined with a DNN was presented for the augmentation of war trauma data. Compared with the traditional artificial data augmentation method, our integrated modeling approach not only improves the quality of injury consequence assessment, but can also automatically generate large-scale and credible virtual war trauma data. The generated data make it possible to carry out related data-based military research, which has great practical significance and value. In addition, it also provides a practical auxiliary tool for quickly evaluating soldiers' injuries and formulating treatment strategies, which are of crucial significance to the analysis and evaluation of war trauma data. Finally, because this study was the first attempt to combine DL and the trauma scoring algorithm for the augmentation of war trauma data, it still had some shortcomings, but with the continuous improvement of the WTSS algorithm, the performance of our WTSS–DNN integrated model will become more superior. That is also the focus and direction of our future research, to continuously improve the comprehensiveness and applicability of our integrated modeling approach.

**Author Contributions:** P.Z. conceived the presented idea and verified the analytical methods. J.Y. provided the experimental environment, supervised and validated the findings of this work. Y.Z. provided the research topic, medical theoretical and technical support. Writing, editing, and formatting the manuscript was carried out by P.Z. with support from Y.Z. and J.Y. Funding acquisition was carried out by S.W. and Y.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Fund, sponsor: Jibin Yin, funding number: 61741206.

Acknowledgments: We would like to thank Shuoyu Wang and Yi Han for their assistance with this study.

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

### References

- Cardi, M.; Ibrahim, K.; Alizai, S. Injury patterns and causes of death in 953 patients with penetrating abdominal war wounds in a civilian independent non-governmental organization hospital in Lashkargah, Afghanistan. World J. Emerg. Surg. 2019, 14, 1–8. [CrossRef] [PubMed]
- 2. Crevier, D. AI: The Tumultuous History of the Search for Artificial Intelligence; Basic Books, Inc.: New York, NY, USA, 1993.
- 3. Pan, Y. Heading toward Artificial Intelligence 2.0. Engineering 2016, 2, 409–413. [CrossRef]
- 4. Zhuang, Y.; Chen, C.; Wu, F.; Pan, Y. Challenges and opportunities: From big data to knowledge in AI 2.0. *Front. Inf. Technol. Electron. Eng.* **2017**, *18*, 3–14. [CrossRef]
- 5. Salem, A.; Moselhi, O. AI-based cloud computing application for smart earthmoving operations. *Can. J. Civ. Eng.* **2021**, *48*, 312–327. [CrossRef]
- Zheng, A.; Chen, Z.; Li, C. Learning Deep RGBT Representations for Robust Person Re-identification. Int. J. Autom. Comput. 2021, 18, 443–456. [CrossRef]
- Liu, J.; Wang, L.; Zhou, H. The Application of Human–Computer Interaction Technology Fused with Artificial Intelligence in Sports Moving Target Detection Education for College Athlete. *Front. Psychol.* 2021, 2848. [CrossRef]
- 8. Tang, Z.; Shao, K.; Zhao, D.; Zhu, Y. Recent progress of deep reinforcement learning: From AlphaGo to AlphaGo Zero. *Control Theory Appl.* **2017**, *34*, 1529–1546. [CrossRef]
- He, T.; Mamta, P.; Richard, O.; James, M.; Yu, X.; Chen, S. Deep learning analytics for diagnostic support of breast cancer disease management. In Proceedings of the IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Orlando, FL, USA, 16–19 February 2017.
- Duan, X.; Yang, Y.; Tan, S.; Wang, S.; Feng, X.; Cui, L.; Feng, F.; Yu, S.; Wang, W.; Wu, Y. Application of artificial neural network model combined with four biomarkers in auxiliary diagnosis of lung cancer. *Med Biol. Eng. Comput.* 2017, 55, 1239–1248. [CrossRef]
- 11. King, P. Signal Processing and Machine Learning for Biomedical Big Data. IEEE Pulse 2019, 10, 34–35. [CrossRef]
- Lee, J.; Jun, S.; Cho, Y.; Lee, H.; Kim, G.; Seo, J.; Kim, N. Deep learning in medical imaging: General overview. *Korean J. Radiol.* 2017, 18, 570–584. [CrossRef]
- 13. Fan, R.; Zhang, N.; Yang, L.; Ke, J.; Zhao, D.; Cui, Q. AI-based prediction for the risk of coronary heart disease among patients with type 2 diabetes mellitus. *Sci. Rep.* **2020**, *10*, 14457. [CrossRef]
- 14. Rong, G.; Mendez, A.; Assi, E.; Zhao, B.; Sawan, M. Artificial Intelligence in Healthcare: Review and Prediction Case Studies. *Engineering* **2020**, *6*, 91–301. [CrossRef]
- 15. Menchón-Lara, R.; Sancho-Gómez, J.; Bueno-Crespo, A. Early-stage atherosclerosis detection using deep learning over carotid ultrasound images. *Appl. Soft Comput.* **2016**, *49*, 616–628. [CrossRef]
- 16. Liu, Y.; Zhou, Y.; Liu, X.; Wang, C.; Wang, Z. Wasserstein GAN-Based Small-Sample Augmentation for New-Generation Artificial Intelligence: A Case Study of Cancer-Staging Data in Biology. *Engineering* **2019**, *5*, 156–163. [CrossRef]
- 17. Ellery, W.; David, F. Deep Learning-Based Survival Prediction for Multiple Cancer Types Using Histopathology Images. *PLoS* ONE. **2020**, *15*, e0233678. [CrossRef]
- 18. Joohi, C.; Puneet, G. BPBSAM: Body part-specific burn severity assessment model. Burns 2020, 46, 1407–1423. [CrossRef]
- 19. Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning; MIT Press: Cambridge, MA, USA, 2016; p. 226.
- 20. Mamoshina, P.; Vieira, A.; Putin, E.; Zhavoronkov, A. Applications of deep learning in biomedicine. *Mol. Pharm.* **2016**, *13*, 1445–1454. [CrossRef]
- 21. Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J. Deep learning for healthcare: Review, opportunities and challenges. *Brief. Bioinform.* **2018**, *19*, 1236–1246. [CrossRef] [PubMed]
- 22. Bozkurt, S.; Gimenez, F.; Burnside, E. Using automatically extracted information from mammography reports for decision-support. *J. Biomed. Inform.* **2016**, *62*, 224–231. [CrossRef]
- 23. DeRouin, E.; Brown, J.; Beck, H.; Fausett, L.; Schneider, M. Neural network training on unequally represented classes. *Intell. Eng. Syst. Artif. Neural Netw.* **1991**, *1*, 135–145.
- 24. Chawla, V.; Bowyer, W.; Hall, L. SMOTE: Synthetic minority over-sampling technique. J. Artif. Intell. Res. 2002, 16, 21–357. [CrossRef]
- Han, H.; Wang, W.; Mao, B. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In Proceedings of the International Conference on Intelligent Computing, Hefei, China, 23–26 August 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 878–887.
- Cieslak, A.; Chawla, V.; Striegel, A. Combating imbalance in network intrusion datasets. In Proceedings of the 2006 IEEE International Conference on Granular Computing, Atlanta, GA, USA, 10–12 May 2006; pp. 732–737.

- He, H.; Bai, Y.; Garcia, A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008; pp. 1322–1328. [CrossRef]
- Barua, S.; Islam, M.; Yao, X.; Murase, K. MWMOTE–Majority weighted minority oversampling technique for imbalanced data set learning. *IEEE Trans. Knowl. Data Eng.* 2012, 26, 405–425. [CrossRef]
- 29. Douzas, G.; Bacao, F. Self-Organizing Map Oversampling (SOMO) for imbalanced data set learning. *Expert Syst. Appl.* **2017**, *82*, 40–52. [CrossRef]
- 30. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *2*, 27.
- 31. Wang, K.; Gou, C.; Duan, Y.; Lin, Y.; Zheng, X.; Wang, F. Generative adversarial networks: Introduction and outlook. *IEEE/CAA J. Autom. Sin.* 2017, *4*, 588–598. [CrossRef]
- 32. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* 2018, 35, 53–65. [CrossRef]
- Canzi, G.; De Ponti, E.; Novelli, G.; Mazzoleni, F.; Chiara, O.; Bozzetti, A.; Sozzi, D. The CFI score: Validation of a new comprehensive severity scoring system for facial injuries. J. Cranio-Maxillofac. Surg. 2019, 47, 377–382. [CrossRef]
- Ratnayake, H.; Johnson, D.; Martensson, J.; Lam, Q.; Bellomo, R. A laboratory-derived early warning score for the prediction of in-hospital mortality, ICU admission, Medical Emergency Team activation and Cardiac Arrest in general medical wards. *Intern. Med. J.* 2019, 37. [CrossRef]
- 35. Trongtrakul, K.; Patumanond, J.; Kongsayreepong, S.; Morakul, S.; Pipanmekaporn, T.; Akaraborworn, O.; Poopipatpab, S. Acute kidney injury risk prediction score for critically-ill surgical patients. *BMC Anesthesiol.* **2020**, *20*, 1–10. [CrossRef]
- 36. Zhou, J. Introduction of Trauma Scoreology. Inj. Med. 2018, 7, 4-9.
- 37. de Alencar Domingues, C.; Coimbra, R.; Sérgio, R.; Poggetti, R.; de Souza Nogueira, L.; de Sousa, R. New Trauma and Injury Severity Score (TRISS) adjustments for survival prediction. *World J. Emerg. Surg.* **2018**, *13*, 12. [CrossRef]
- 38. Yang, J.; Liu, Z.; Zhang, Y. Comparison of injury degree score method and modified trauma severity score method for inpatients with accidental injury. *Med. J. Natl. Defending Forces Northwest China* **2017**, *38*, 364–368.
- 39. Shi, J.; Shen, J.; Zhu, M.; Wheeler, K.; Lu, B.; Kenney, B.; Xiang, H. A new weighted injury severity scoring system: Better predictive power for adult trauma mortality. *Inj. Epidemiol.* **2019**, *6*, 40. [CrossRef] [PubMed]
- 40. Wang, H.; Liu, X.; Wang, G. Application and analysis of revised injury severity score on emergencies. *Orthop. Biomech. Mater. Clin. Study* **2014**, *11*, 20–22.
- 41. Yang, T.; Guo, P.; Liu, W.; Liu, X.; Hao, T. Enhancing PIR-based Multi-person Localization through Combining Deep Learning with Domain Knowledge. *IEEE Sens. J.* 2020, 1. [CrossRef]
- 42. Ding, L.; He, H. Research on domain adaptation of machine translation based on domain knowledge and deep learning. *Inf. Sci.* **2017**, *35*, 125–132.
- 43. Mcdaniel, P.; Papernot, N.; Celik, B. Machine learning in adversarial settings. IEEE Secur. Priv. 2016, 14, 68–72. [CrossRef]
- 44. Gordon, J. The delphi method. Futures Res. Methodol. 1994, 2, 1-30.
- 45. Penn-Barwell, J.G.; Bishop, R.; Midwinter, J. Refining the Trauma and Injury Severity Score (TRISS) to Measure the Performance of the UK Combat Casualty Care System. *Mil. Med.* **2018**, *183*, e442–e447. [CrossRef]
- 46. Liu, Y.; Xu, Z.; He, J.; Wang, Q. Data Augmentation Method for Power Transformer Fault Diagnosis Based on Conditional Wasserstein Generative Adversarial Network. *Power Syst. Technol.* **2020**, *44*, 1505–1513.
- 47. Diederik, K. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
- 48. Zhang, H.; Huang, L.; Wu, C.; Li, Z. An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset. *Comput. Netw.* **2020**, 177. [CrossRef]
- Xu, R.; Cao, J.; Wu, Y.; Wang, S.; Luo, J.; Chen, X.; Fang, F. An integrated approach based on virtual data augmentation and deep neural networks modeling for VFA production prediction in anaerobic fermentation process. *Water Res.* 2020, 184, 116103. [CrossRef] [PubMed]
- 50. Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
- 52. Webb, G. Naïve Bayes. Encycl. Mach. Learn. 2010, 15, 713–714.