An Intelligent Model for Predicting the Occurrence of Skiing Injuries

Fisnik Dalipi, Diana Marina Armijo Mendoza, Ali Shariq Imran, Sule Yildirim Yayilgan
Gjøvik University College
Faculty of Computer Science and Media Technology
Postboks 191, 2802 Gjvik, Norway
Email: {fisnik.dalipi, diana.mendoza, ali.imran, sule.yayilgan}@hig.no

***Abstract*—Artificial neural networks offer a unique way to model very complex and innovative systems that can be very effective in anticipating various accident severities. In this article, we propose a neural-network-based model, able to predict the number of severe injuries caused while skiing. The proposed system is intended for use by ski patrol and medical personnel to better prepare themselves in advance for treating ski-injured persons. The ski patrol and any other medical personnel will be able to know the statistics, type and severity of the injuries occurred, and most importantly, will be benefiting from having predictions for each day. Considering that, the number of injured people in a particular place each day was estimated, the results are very promising suggesting that such a system would prove beneficial in accurately predicting skiing injuries.**

I. Introduction

Skiing is a winter sport that gains popularity every year. As in any other activity, it is necessary to follow some rules in order to ensure our and other's safety. However, all the precautions taken does not necessarily provide the safety we require. Consequently, new technologies have become useful instruments for skiers in providing the required safety. One example of these technologies is the RECCO®System [1]. This is a small reflector included in ski clothes, such as coats that echoes the signal of radar carried by ski patrol or other rescue teams when an avalanche has buried someone. Some groups like the Internationals Ski Federation (ISF) [2], have been researching, gathering data and developing a medical guide [3] to prevent injuries or reduce them in the ski disciplines at the elite level. As said in the medical guide, the guide is not intended as a book of medicine specialties, but to help manage possible injuries when they occur and in the first place to prevent them. For example, there is a complete explanation of the communication that has to be given to the medical personnel during the transportation of the injured person, and the data that must be gathered about the injured person, such as the personal information, specification of the injury, weather conditions and a video recording of the injury. Some other devices and applications have been developed to warn of an accident to the involved authorities, such as snow Penetrometer [4], but there are only a few to prevent or predict injuries. Recently, one study has been made by the Department of Orthopedics and Traumatology in Helsinki University Hospital, Finland. A system called SKIDATA®has been developed and it records automatically every ski-lift run taking place in Levi Ski Resort [5]. This system has been taking information over six years maintaining a record of the average number of injury incidences, and a complete study of the most frequent injuries caused while skiing. The injuries are classified in different severities, being Grade 1 as minor and Grade 4 as critical. Having all that information, it was demonstrated that the most common injury among skiers is on the lower and upper extremity. This can be seen in Figure 1.

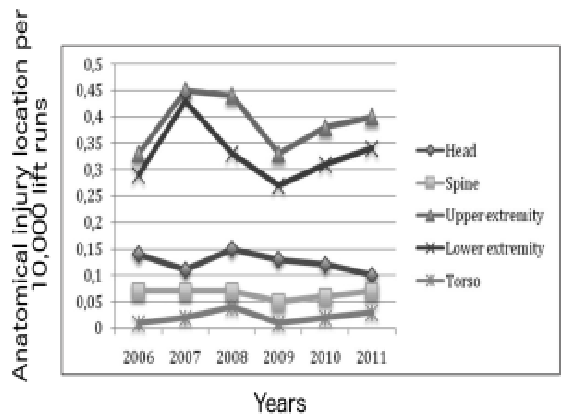

Fig. 1. Frequent injuries recorded with SKIDATA®

The work made by the Helsinki University Hospital gathers an important information that can be used, as they proposed, to discuss different preventive measures to reduce the number of injuries caused in the future. The rest of the paper is structured as follows: Section II describes the Project Description. Section III describes the parameters for risk factor estimation while section IV presents the performance evaluation results. Lastly, section V concludes the paper and gives some possible future directions.

II. Project description

In this paper, we propose a neural-network-based model, to provide prediction capability and intelligence to a digital ski injury registration system in development at our lab, which can be used by the ski patrols to medical unit's chain. By the use of this system, the ski patrol will be able to provide some useful information to the doctors at the hospital in advance and in a timely manner. The ski patrol [6] is an emergency service that looks after people who is injured while skiing, snowboarding or similar. To accomplish their emergency services in the best possible way, they have routines that have been used through the years. Mostly, ski patrols rely on paper-work for keeping records, however by using the appropriate digital tools, record keeping will be easier, and at the same time, the injured person specific data would become more accurate. Nowadays, the process followed by the ski patrol is as follows:

- The ski patrol either gets a call from the injured person or the patrol finds them in the mountain during his/her routine trips.
- The ski patrol fills a paper form with the information provided by the injured person and based on the condition of the injured person. The ski patrol sends the form to the doctor along with the injured person.
- When the injured person arrives to see the doctor's office, the doctor directly examines the injured person without looking at the paper form filled by the ski patrol. Meantime the injured person explains what has happened to the doctor again and gives information about his/her injury.

All this process becomes long and exhausting for the injured person, as well as the medical personnel while the information from the ski patrol goes unnoticed. The overall scenario of our proposed model for predicting ski injuries is shown in Figure 2. Our project, named SkiPatrol, goes beyond proposing, as well as creating a database with the same purpose, the creation of a better way of communication from the doctor to the ski patrol, developing a common application that is useful to all the medical personnel and for improving the injured person care.

As can be seen from Figure 2, once we have the needed information about the injured persons, which are registered to the system's database by the ski patrol, as a next step, we put to the network as many inputs and outputs as we needed, followed by the selection of number of hidden layer and neurons and their activation function. We trained the network with the selection of neurons and finally used the ANN (Artificial Neural Network) with real data to predict the number of injuries while skiing.

The application, among others, will speed up the communication and will have following benefits:

- The injured person will receive the best attention/care, since the doctor has already studied the case in advance.
- The doctor will receive injured persons' status information in advance. Thereby, the medical personnel will obtain the necessary time to prepare themselves with the best treatment for the incoming person and to have the entire equipment ready on time when needed.
- All the information gathered from the injured person in the ski resort will be stored and transmitted confidentially from the first moment with digital security. There will be a complete, updated and reliable database where information such as i) Injured person information: ID, name, nationality, birthday, gender, weight, height, etc. is stored. This information might help the doctor to select the best treatment for the injured person. Moreover, having the information would make it easier to integrate incident related information on the existent EPJ (Electronic Patient Journal); ii) Incident: Date of registration and injury, time of the injury, sports type (ski, snowboard, slalom, etc.), kind of equipment (borrowed, rented, or new); iii) Injury type: body part damaged, type of injury (concussion, contusion, fracture, etc.) and severity, each injury identified with a different number into the digital system, the ski patrol will receive feedback in the form of reminders to attend the injured person better. For example, if the injured person got hurt in the neck, the ski patrol will receive a reminder to use a neck brace.

This model will provide the possibility of being applied in any country by changing some variables and therefore being able to enlarge and optimize the database.

A. *Overview of the proposed ANN*

Artificial Neural Networks were first introduced by McCulloch & Pitts in 1943 [7], with their idea of simplified neurons. Nevertheless, until some important theories, like error back-propagation, were proposed in the early eighties, and the hardware developments increased the processing capacities, the interest in this field did not grow. ANN are formed with a pack of simulated neurons that communicate with each other over weighted connections, like the synapses of the human brain. When a neuron receives an input, it makes a small operation with it, and sends the result to the next one. Neural Networks are ideal in recognizing diseases using scans since there is no need to provide a specific algorithm on how to identify the disease. Neural networks learn by training so the detail of how to recognize the disease is not needed. There are several machine-learning algorithms to develop an ANN [8]. In our project, we are using Supervised Learning (Machine Learning). This kind of algorithm uses a known dataset to make predictions, which includes input data and known response values for that data. The Supervised Learning algorithm seeks to build a model that can make predictions of the response values for a new dataset. Once that model is calculated, it can be tested with new data. Training with large datasets often yields models with higher predictive power that can generalize well for new datasets. We created the artificial neural network using Matlab Neural Network toolbox. The Neural Network is a Feed-Forward network with tan-sigmoid transmission function, used as the activation function in the hidden layers. The training data is a combination of values of variables and is fed as input to the neural network for training given the respective outputs [9]. The dataset contained 181 inputs, one for each day of 2013-2014 ski season from Trysil ski station, one of the largest Norway's ski destination. The following values were considered in the analysis:

- Input variables: day of the year (from the 1^{st} of November to the 30^{th} of April, i.e. from 305 to 365 and from 1 to 120) day of the month, day of the week, holidays,

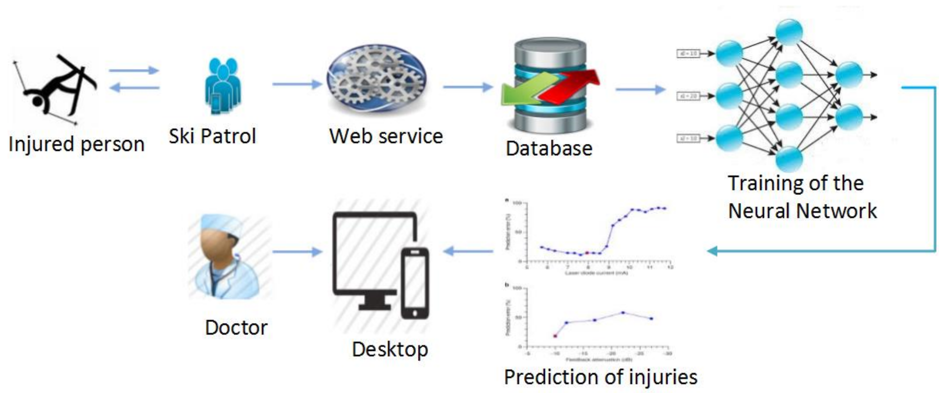

Fig. 2. Injury prediction workflow scenario

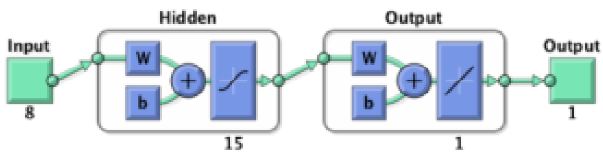

Fig. 3. The Feed-Forward Neural Network architecture

minimum temperature, foreseen affluence, snow depth and precipitation.
- Targeted: Estimated people mild-severe injured.

The topology of the trained network is given in Figure 3. Regarding to the network configuration, we use 15 neurons for each training, where 70 % of the input values as training samples, 15 % as validation samples and the rest as test samples.

The training samples are the ones used for training the network. Training needs a large amount of samples because that is the source of knowledge for the ANN, where it adjusts the error based on the given input-output mappings in the training set. Validation samples are used to measure network's generalization capability, and to halt training when generalization stops improving. The test samples have no effect on the training and so provide an independent measure of the network performance during and after the training [10].

III. Risk factor estimation

To estimate the injured people, we calculated the possible risks of each factor that can result in people falling and getting injured. All these estimations are selected according to the next premises:

- Day risk: It is assumed to be 3 if it is holiday, and will also increase according to Table 1. It is observed that during the weekdays the risk is low, because normally people go skiing during the weekend and holidays. The day risk, including holidays, will be guessed to be between 0 and 3.
- Affluence risk: The more people, the more risk. This is because if skiers do not have enough space, they can collide with each other. In addition, there might be many inexperienced skiers that can fall more often. This value will vary from 0 to 4, and follows Table II, knowing the fact that there can be a maximum capacity
- Snow depth associated risk: If there is not enough snow and the ski station is open, for some reason people can go out of the lane and fall because of the rocks emerging out of the snow. We assume that, if the snow depth is less than 20 cm, the ski station is not opened, and therefore, there is no risk. If the snow depth is between 20 and 50 cm, the station is opened, which is assumed as an intermediate risk for skiers. If the snow depth is bigger, there will be low risk. The values of snow risk are given in Table III.
- Temperature risk: In order to calculate the temperature risk, we need to relate this risk with the skiing conditions. In the cases when the temperature is too low and it is not snowing, the snow can freeze, hence skiing in these conditions can be difficult. People have more probabilities

TABLE I
Day risk

Day of the week	Risk
Monday,Tuesday, Wednesday, Thursday	0
Friday	1
Saturday	3
Sunday	2

TABLE II
Affluence risk

Foreseen affluence (FA) in number of people	Affluence Risk (0-3)
0 <FA <1500	0
1500 <FA <4500	1
4500 <FA <7000	2
7000 <FA <9000	3

TABLE III
Snow risk

Snow depth (SD) in cm	Snow Risk
0 <SD <20	0
20 <SD <50	2
SD >50	2

TABLE IV
TEMPERATURE RISK

Temperature [T] in 0^0C	Temperature Risk
T >0	0
-10 <T <0	1
T <-10	2

TABLE V
PRECIPITATION RISK

Precipitation [P] in mm	Precipitation Risk
P <10	0
10 <P <20	1
T >20	3

to slide and be injured. It is assumed that temperatures lower than 0^0C will increase the risk, because ice can appear on top of the snow, making it difficult to ski. Temperatures lower than 10^0C are assumed to increase the risk. These variables can be easily adjusted according to the domain knowledge of the ski station specialists.

- Precipitation risk: If it is snowing or there is fog, visibility is poor, and this can increase the risk. In this case, it is considered that a precipitation between 10 and 20 mm can disturb the skiers, and more than 20 mm, is assumed more risky for them.
- Total day risk: This is a value that sums up all the assumed risks. It will have values from 0 to 13 and is the addition of previous risks explained.
- Estimated People Injured [EPI]: This estimation will be calculated using the formula in 1.
- Estimated people mild-severe injured: This will be the 1 % of the EPI calculated before.

$$EPI = 0.01 * TotalDayRisk * ForeseenAffluencePeople \quad (1)$$

IV. PERFORMANCE EVALUATION

In order to know the amount of needed hidden neurons, we made some tests to find the best results. Finally, it was decided to use 15 neurons, using the Levenberg-Marquardt back propagation algorithm. It is important to note that for recognizing the network efficiency we use Mean Squared Error (MSE), which is the average squared difference of the samples between outputs results from the ANN and the targets introduced in the dataset. Zero means no error and, therefore, smaller values are better. From Table 6 we see that MSE for the training samples is 0.028 injured people, but it has smaller values or shows better results on testing samples, which is relevant to see how the network works with real values. All the values of regression (R) for this test are very close to 1, but the testing have smaller values. Nevertheless, the obtained results are promising.

Figure 4 shows the Mean Squared Error (MSE) vs. epochs or iterations. In the figure, we can see that the fourth iteration was the one at which the validation performance reached the minimum, and it has the smaller error. The training continues

TABLE VI
MSE AND REGRESSION VALUES FOR THE TRAINING, VALIDATION AND TESTING

	Samples	MSE	R
Training	127	2.78052e-2	9.89058e-1
Validation	27	8.76445e-2	9.7087e-1
Testing	27	1.08357e-1	9.42303e-1

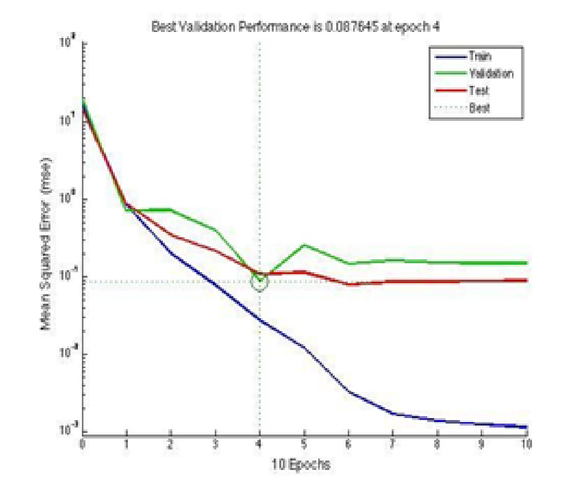

Fig. 4. Diagram of training, validation and testing error

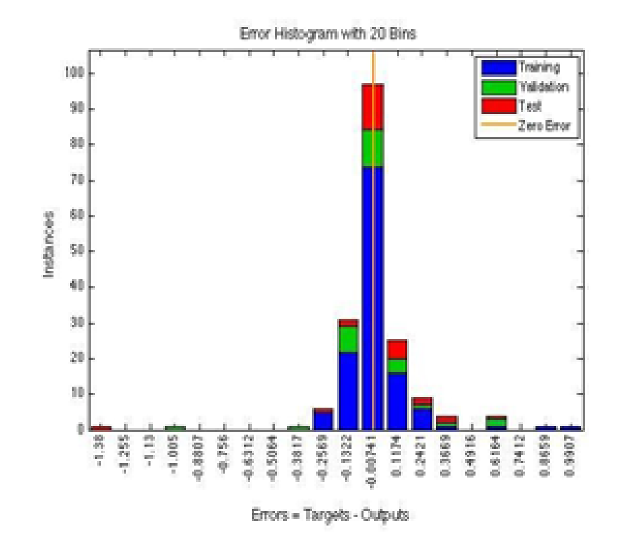

Fig. 5. Histogram of errors

6 more epochs (iterations) before the training stopped. In the same figure, we see that the test curve has no over fitting, as it has not increased significantly before the validation curve increased. In fact, the curve remains stable in the next iteration and then it slightly decreases.

The histogram in Figure 5 presents the number of instances (inputs) per error. The error is measured with the subtraction of the target introduced before and the resultant output from the ANN. We see that the biggest error in this training was of one person (1.38 person) on approximately one test instance.

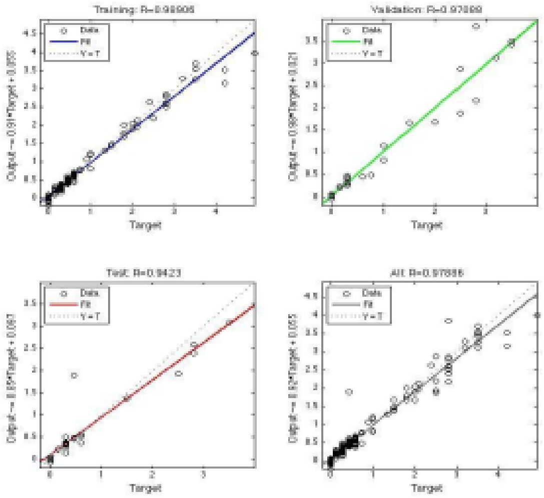

Fig. 6. Diagram of all outputs

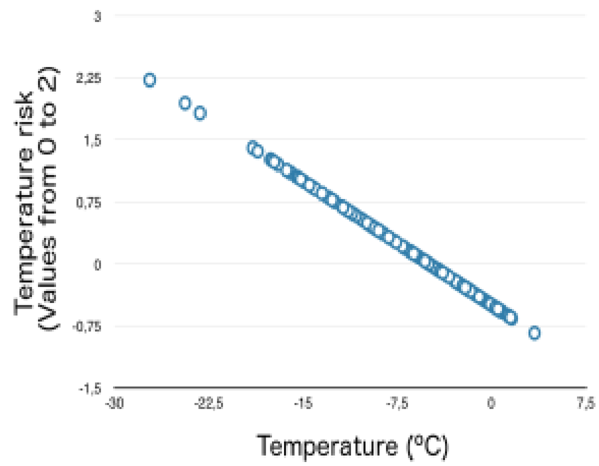

Fig. 7. Lineal estimation of temperature risk

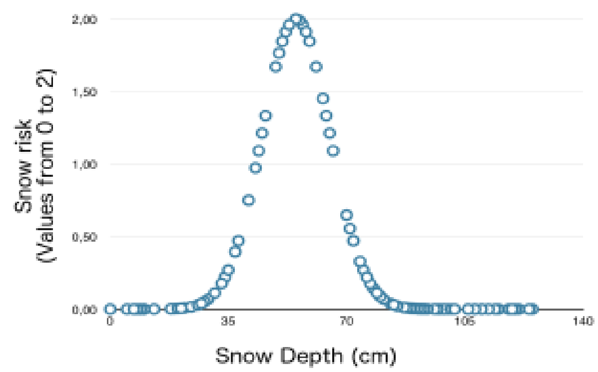

Fig. 8. Gaussian curve of snow risk

The last figure, Figure 6, shows four plots representing training, validation, test and the addition of all of the outputs. Each plot represents the linear regression of targets relative to outputs [11]. The solid line in each of them shows the perfect result of the ANN result minus the output entered.

In order to have a smaller error, and even better correlation between outputs and targets (R), we decided to improve the dataset. To make this, we improved the input data by making lineal and Gaussian estimations of the different risks instead of taking discrete values. The changed data was the temperature risk, and the snow risk, leaving the affluence risk discreet. This data still have the same premises that had, but they vary with continuous values. The curves of this new data are shown in Figure 7 and Figure 8.

Figure 7 presents a lineal solution that increases the risk having the biggest risk around -30, of approximately 2, and the smaller around 0 with 0 risk. The formula of this graph is the following:

$$TemperatureRisk = -0.5 - minTemperature * 0.1 \quad (2)$$

Figure 8 represents the same table (Table III) explained above but with a Gaussian curve, being values under 20 and above 80 cm of depth approximately of low risk and the biggest risk around 40 cm, because there is not enough snow, but the ski station is opened. Gaussian formula is given by equation 3

$$f(x) = a * exp^{(-\frac{(x-b)^2}{2*c^2})} + d \quad (3)$$

We define f(x) as the Snow risk, x as the snow depth in centimetres for each day, and the variables a, b, c, d as following: a=2, b=55, c=10 and d=0. The configuration of the network is the same as with the first test, having 15 hidden neurons, 70 % of the samples for training and the rest divided equally between test and validation samples. The results with these changes were promising, as can be seen in Table VII. This table shows a very small MSE for the three kinds of sample data, having a small value of mean squared error of 0.003 (3 people out of 1 thousand of error) on the test data, and a very good results for the correlated data, almost correlated 100 % in all the samples taken.

TABLE VII
MSE AND REGRESSION VALUES FOR THE TRAINING, VALIDATION AND TESTING WITH IMPROVED DATASET

	Samples	MSE	R
Training	127	1.07387e-3	9.99336e-1
Validation	27	1.18499e-2	9.94836e-1
Testing	27	3.06513e-3	9.94659e-1

The figures in relation to these results are given in Figure 9 and Figure 10, where in Figure 9 it is demonstrated that there is no over fitting and the curves stabilize, in the case of validation and test, around the 6^{th} epoch.

The histogram in Figure 10 shows that the error is of around 0.5 person, that rounding is zero.

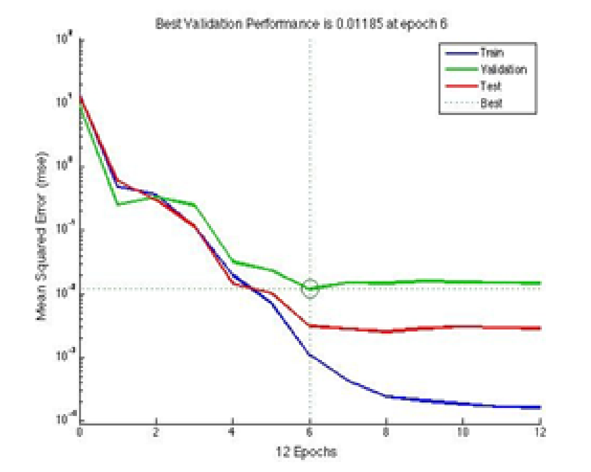

Fig. 9. Diagram of training, validation and testing error with improved

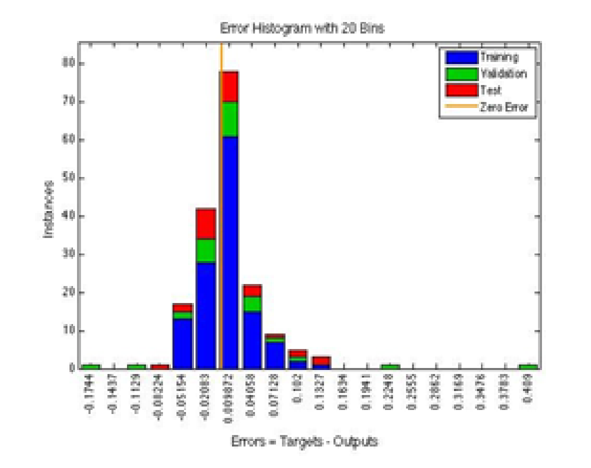

Fig. 10. Histogram of errors with improved dataset

Finally, in Figure 11, the plot shows that the correlation between outputs and targets is much better, compared to the previous dataset.

V. Conclusions

This work proved that it is possible to predict the number of injured people in a particular place. Although some of the data are estimated, the results are valuable, because the input data can be changed as soon as more data is provided. The errors we obtained in the last testing were of a MSE of 0.003. This means that for each 1000 people, the trained ANN will have an error of 3 people. Considering that the number of injured people each day was estimated, the result values are excellent, determining that it is possible to predict injured people using this algorithm, but the results will be improved when more data will be used. As a future work to be considered is the introduction of additional relevant data, such as actual risk values, and GPS coordinates of the place the persons were injured. This will provide more accurate output information and new information, such as a map with the most risky areas, avalanche predictions and other natural disasters.

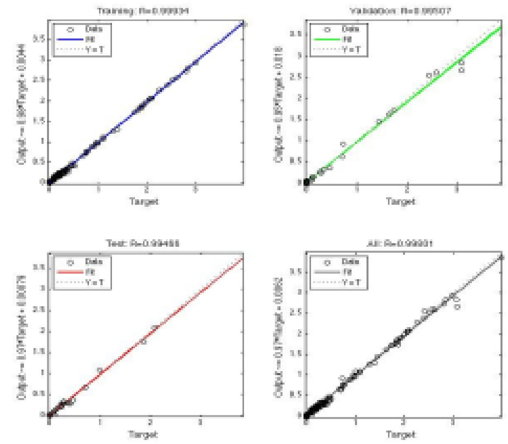

Fig. 11. Diagram of all outputs with improved dataset

References

[1] RECCO, The RECCO System [Internet]. 2013 [cited 2014 September] Available from: http://www.recco.com/the-recco-system

[2] Oslo Sports Trauma Research Center, FIS and Oslo Sports Trauma Research Centre [Internet]. 2014 [cited 2014 September]. Available from: http://www.klokavskade.no/en/News-archive/News-archive-2006/FIS-MEDIA-INFO

[3] International Ski Federation, FIS Medical Guide [Internet]. 2011 [updated 2013, cited 2014 September]. Available from://www.fisski.com/mm/Document/documentlibrary/Medical/03/31/99/fis-medical-guide-2013$_Neutral$.pdf

[4] NewScientist: Snow penetrometer is developed by the Swiss Federal Institute for Snow and Avalanche Research in Davos [Internet]. 2014 [cited 2014 September] Available from: http://www.newscientist.com/article/dn3410-penetrometer-predicts-risk-of-avalanches.html.VDj4URCbH0N

[5] A. Stenroos, L. Handolin, Incidence of recreational alpine skiing and snowboarding injuries: six years experience in the largest ski resort in Finland, US National Library of Medicine National Institutes of Health, PubMed 2014

[6] SkiStar Trysil. Mountain Safety [Internet].2014 [cited 2014 August]. Available from: http://www.skistar.com/en/trysil/skipass-skiing/mountain-safety.

[7] S. Warren, McCulloch and Walter Pitts, A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics. Volume 5, 1943.

[8] R.P. Lippman, An introduction to computing with neural nets. IEEE ASSP Magazine, 1987.

[9] M.T. Hagan, H.B. Demuth, and M.H. Beale, Neural Network Design, Boston, MA: PWS Publishing, 1996.

[10] Matlab Documentation Centre. Neural Network Toolbox Bibliography [Internet]. 2014 [cited 2014 September] Available from: http://www.mathworks.es/es/help/nnet/ug/neuralnetwork- toolbox-bibliography.html.

[11] The Mathworks, Inc. Plotregression, [Internet]. [cited 2014 September]. Available from: http://www.mathworks.se/help/nnet/ref/plotregression.html