Identifying ski roughness using data driven approaches

Bassma Al-Jubouri[1,2] and Astrid S. de Wijn[1,3]

***Abstract*— In skiing sport, snow friction is a crucial factor in determining the ski roughness that can produce high speed and quick finishing time for a skier. However, snow friction is influenced by many factors associated with weather and snow conditions which affect the choice of the optimal ski's roughness. This paper proposes an ensemble learning system that can accurately recommend the best ski roughness under different weather conditions. The data used in this study is a unique data set that has been collected from field tests and competitions. Though this data set recorded information about ski treatment and weather conditions over a 10-years period, it is affected by noise and outliers, and it has an imbalanced distribution in the ski roughness classes. This work addresses these challenges in the data by applying pre-processing techniques and class balancing strategies. Furthermore, correlation and clustering approaches are employed to identify redundancies in the data and to recognise the subsets of weather conditions that have the highest influence on the selection of the ski roughness. Using the resultant clusters, an ensemble system is introduced to recommend the most suitable skis roughness for a given weather condition. This system can be used as a guiding tool in skiing competitions to aid technicians in choosing the skis roughness. The results showed that air and snow temperatures as well as snow humidity have the highest impact on the choice of the ski roughness.**

I. INTRODUCTION

In winter sports, most snow friction studies are conducted in a well-controlled lab environment and the results are then employed in real world competitions [1], [2], [3]. Typically, these studies focus on ice friction, because ice surfaces can be reproduced to a pre-defined specification relatively easily compared to the more complex structure of snow [1]. The results obtained from the ice surface experiments are then generalized to snow. This generalization often causes a gap between the theoretical findings and the actual performance [4]. This study aims to bridge this gap using a large data set that has been collected from professional ski technicians preparing skis for real world competitions over a period of 10 years.

Machine learning approaches have been utilized in snow and ice friction studies. In particular, those related to road friction and safety driving, such as: the use of deep Convolutional Neural Network (CNN) to classify road surfaces into different categories to estimate the friction [5] and the use of support vector machine to classify winter road surface conditions images obtained from low cost cameras [6].

In winter sports, machine learning has been applied to identify skiing strategies, such as predicting cross-country skiing techniques [7], and identifying alpine skier posture [8]. However, there have been few studies that examine the effect of snow and weather conditions on the ski-snow friction using real field test data sets. One example for these studies is presented in [4] where 175 field tests were recorded and analyzed using correlation and regression methods.

In this paper, we use a unique large data set collected from real world field tests and competitions. This data set was provided by Olympiatoppen [9]. Using this data, we demonstrate the influence of different snow types and weather conditions on the snow friction through the choice of the roughness of the skis. Furthermore, this paper examines the relation between snow and weather conditions to identify the similarities and differences between them using correlation and clustering approaches. This work introduces an ensemble model that can accurately predict the ski roughness required to

*This work was not supported by any organization

[1]Author is with Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway.

[2]bassma.al-jubouri@ntnu.no

[3]astrid.dewijn@ntnu.no

minimize the snow friction and increase the skier speed. Furthermore, we identify the most important weather features in winter sport that control the choice of the best ski roughness.

The remainder of this paper is organised as follows: Section II provides an overview of the main challenges associated with the skiing data set and describes the data features and classes. Section III discusses the data preparation and examines the dependencies among the weather conditions using correlation and clustering approaches. Section IV introduces the methodology of the experimental work and explains the ensemble model used to predict the ski roughness. The results and discussion are presented in Section V. Finally, Section VI draws the main conclusions and outlines the possible directions of the future work.

II. DESCRIPTION OF THE DATA SET

The data set used in this study was provided by Olympiatoppen [9], which is part of the Norwegian Olympic and Paralympic committee. It contains a series of field tests carried out between 2009 and 2019, and it has 10074 data instances with 64 features. The data set features describe the performance of different ski roughness profiles under a variety of snow and weather conditions.

Up to the author's knowledge, this is the first data set that presents a large amount of field tests that spans such a long period of time and covers a wide range of weather conditions and ski treatments. As mentioned before, most studies found in literature focus on data obtained from a series of controlled-lab tests. Thus, using the data set presented in this study has the potential to bridge the gap between laboratory findings and actual performance in real world competitions. Furthermore, the results obtained using this data can be extended to other fields, where such detailed information for snow friction is not available. For example, in road safety where the friction coefficient is often estimated using simple road classification into bare, snow, or icy road conditions [5] [10] [6].

There are several challenges associated with this data set. First, it has an unbalanced class distribution in terms of the tested roughness profiles (i.e. the grind classes). In addition, it suffers from inconsistency, as many parameters were added and removed over time. Another challenge is noise in the recorded values for some parameters, which is mainly caused by the methods and tools used to record them, as well as the human error factor.

This study focuses on the main nine features that describe the weather and snow conditions when the skiing tests were performed, and their effect on the choice of the ski roughness. These features are:

- Air temperature and snow temperature: measured in degrees Celsius.
- Relative air humidity and snow humidity: measures the water content in the air and in the snow as a percentage in the ranges $0\% - 100\%$.
- Snow hardness: measures 7 levels of snow hardness ranging from: very low, low, middle, middle-high, high, very high and ice.
- Snow grain size: measures the diameter of the snow grains in mm.
- Precipitation. This feature has five possible values: no rain, light rain, rain, light snow, and snow.
- Natural and artificial snow: these two features refer to the type of snow, and can have the following values: fallen, new, old, converted, mixed and salted.

We investigate 10 classes of grinds ranging from fine grinds to coarse grinds. The roughness profile (measured in micrometers (μm)) is classified into the following three categories: fine grind ($< 3\mu$m; classes 3, 9, 10), middle grind (between 3 and 5μm; classes 1, 4, 6, 8) and coarse grind (larger than 5μm; classes 2, 5, 7).

The following section discusses the techniques used to prepare the data and examines the dependencies among the weather conditions using correlation and clustering approaches

III. DATA PREPARATION AND CLUSTERING

In order to address the challenges in the ski data, data preparation and pre-processing steps are carried out. First, all categorical values of the features are mapped into numerical values, using an appropriate scale defined for the individual features. Then all features are normalised to have zero mean and unit standard deviation. The data is then filtered to focus on ten distinguished grind classes that exhibit different roughness profiles.

The imbalanced class distribution in the data is dealt with using oversampling. New data instances are artificially generated by taking the minimum and maximum values of four out of the nine features and producing random values within these limits. The four features used are snow temperature, air temperature, snow humidity, and air humidity. The reason for choosing these four features is that, if a certain grind is used within an upper and lower temperature or humidity limits, then it can be assumed that this grind is used for all the values within these limits. For example, a grind that is used in a minimum temperature of $-15°$C and a maximum temperature of $-5°$C, can be used in the temperature values within the range ($-15°$C,$-5°$C). The same rule applies for humidity values. Meanwhile, the values for the snow type, hardness, grain size and precipitation are kept unchanged, since they define the snow surface conditions.

A. *Correlation between features*

To study the possible dependencies among the features and to detect any redundancies, we compute the correlation coefficients among the nine features. These are shown in Table I.

Three trends in the features correlation coefficients can be seen in Table I, namely:

- There is a high correlation between the air and snow temperatures (0.82). Furthermore, compared to the remaining features, snow humidity has the highest correlation with these two features, of 0.47 and 0.36 respectively.
- The correlation between features that describe snow conditions, including snow hardness, snow grain size, artificial and natural snow types, is moderate, ranging from 0.41 to 0.24. Meanwhile, these features have lower correlation with the other features.
- There is a subset of the features that have low correlation coefficients with the remaining features. These include air humidity and precipitation. It should be noted that the precipitation feature in this data set has the most missing values among the other features.

These three trends are reflected in the feature clustering discussed in the next subsection.

B. *Clustering*

Clustering is an important tool in recognising structures and patterns in the data and in extracting useful information from large data sets [11]. The aim of cluster analysis is to divide N data samples into C homogeneous clusters, such that similar data are grouped into the same cluster.

This work uses agglomerative hierarchical clustering to identify the similarities and differences among the nine features for the weather and snow conditions. Fig. 1 shows the resulting tree. Using a reasonable cut off point, three clusters can be distinguished. The first cluster includes air and snow temperatures as well as snow humidity. Meanwhile, the second cluster includes features that describe the snow conditions (snow grain size, snow hardness and the type of artificial or natural snow). The last cluster groups together precipitation and air humidity. These results show similar patterns as the correlation coefficients discussed above.

The results of the agglomerative hierarchical clustering will be used in Section V to explore how these clusters influence the accuracy of choosing the ski roughness.

IV. METHODOLOGY

The methodology used in this work is illustrated in Fig.2. In this study, the data is pre-processed using the steps discussed in Section III. These steps are: a) changing the categorical values into numerical values, b) normalisation, and c) balancing the class distribution.

After preparing the data, two approaches are implemented to train the ensemble learning model to predict the ski grind as well as to test which features contribute the most to the accuracy of the prediction. Both the overall accuracy of the prediction and the individual class accuracies are measured using a confusion matrix. The first approach uses all the nine available features to train the ensemble learner model. Meanwhile, the second

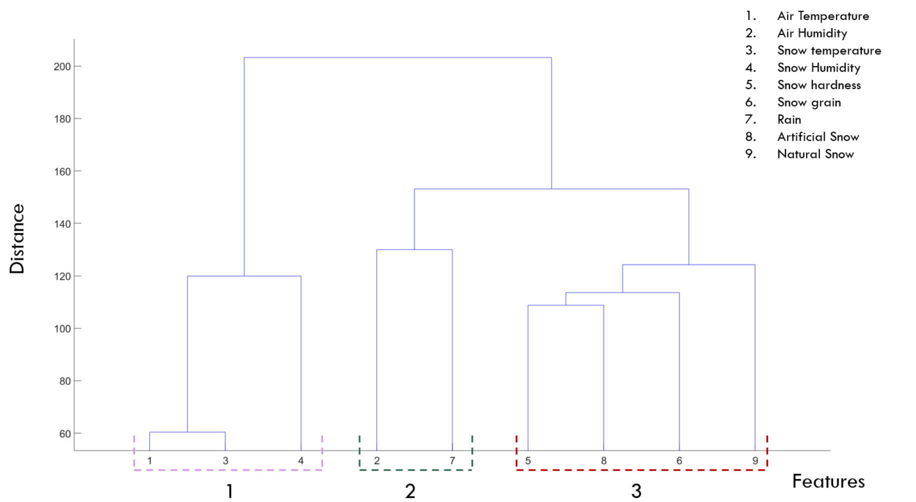

Fig. 1: The hierarchical, binary cluster tree for the nine weather conditions using agglomerative hierarchical clustering method, where the y-axis is the Euclidean distance and the x-axis is the feature number ranging from 1-9.

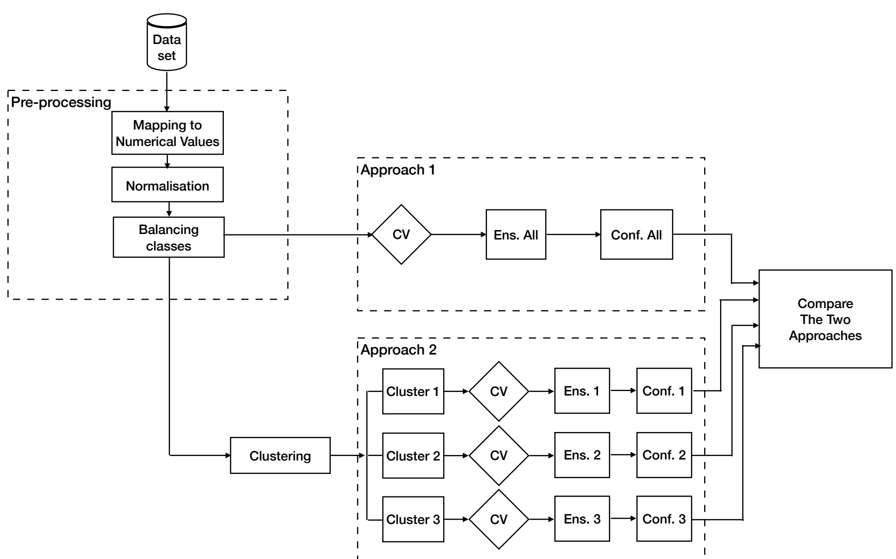

Fig. 2: An illustrative figure for the methodology followed in the two approaches presented in this paper. Here Ens. refers to Ensemble learning, Conf. to confusion matrix and CV to cross validation.

TABLE I: Correlation among the weather features., where AT is Air Temperature, AH is Air Humidity, ST is Snow Temperature, SH is Snow Humidity, SHa is Snow Hardness, SGS is Snow Grain Size, P is Precipitations, AS is Artificial Snow and NS is Natural Snow.

	AT	AH	ST	SH	SHa	SGS	P	AS	NS
AT	1	−0.18	0.82	0.47	−0.19	0.15	−0.09	−0.03	0.09
AH	−0.18	1	0.02	0.11	0.07	0.23	0.16	0.14	0.07
ST	0.82	0.20	1	0.37	−0.27	0.08	−0.09	−0.07	−0.02
SH	0.47	0.11	0.37	1	−0.03	0.30	−0.02	0.12	0.22
SHa	−0.19	0.07	−0.27	−0.03	1	0.33	0.11	0.42	0.24
SGS	0.15	0.23	0.09	0.30	0.33	1	0.05	0.42	0.32
P	−0.09	0.16	−0.09	−0.02	0.11	0.05	1	0.21	−0.04
AS	−0.03	0.12	−0.07	0.12	0.41	0.41	0.210	1	0.29
NS	0.09	0.08	−0.022	0.22	0.24	0.32	−0.04	0.29	1

approach uses the three clusters features described in section III to train the ensemble learner. This approach tests which subset of features has the highest contribution to improving the accuracy of the prediction.

A. *Pool of competing ensemble learners*

Ensembles are predictive systems that combine multiple weak learners to provide the final prediction. Generally, there are two conditions for an ensemble learner to perform better than a single learner. First, the weak learners should be diverse (their error correlation is reduced) and they should have an accuracy better than random guessing [12]. In order to encourage diversity among the ensemble weak learners, they are trained either on randomly selected subsets of the data or on local regions. Data locality can be measured using similarity metrics, such as pairwise squared correlation [13] and conditional mutual information [14].

A pool of competing ensemble learners is used to predict the grind class in the two approaches presented in Fig. 2. Ensembles in this pool are trained to minimise a five-fold cross validation loss function over multiple iterations. The ensemble that has the minimum estimated cross validation error is chosen. The parameters of the system are optimised using Bayesian optimisation. The problem investigated in this study is a multi-class classification problem. Thus, the pool of competing ensembles includes the following multi-class classification ensembles: Bagging [15], AdaBoostM2 [16], LPBoost [17], TotalBoost [17], RUSBoost [18] and Subspace [19]. The weak learners used in the ensembles are random forest decision trees.

V. Results and Discussion

In this section we present the results obtained from applying the two approaches discussed in the previous section. Furthermore, we compare their performance to identify the most important features that control the selection of the best grind class. As mentioned in Section II, the data set used in this work has nine features and ten grind classes. The features describe the weather and snow conditions, and the grind classes describe the roughness of the ski surface. The grind classes vary from fine to coarse grinds.

A. *The first approach: using all features*

In the first approach, all nine features that describe the snow and weather conditions are used to train the ensemble model. A four-fold cross validation is applied to resample the data set, where each time a different fold is used for testing and the remaining three folds are used for training. The cross validation accuracy of the test set, as well as the individual test accuracies of the ten grind classes are illustrated in the confusion matrix shown in Fig. 3a.

Applying this approach resulted in an averaged cross validation error for the training set equals (6.83%) and for the testing set equals (12.99%). The increase in the testing error suggests a slight overfitting. This overfitting can be explained when observing the individual grind classes test accuracies. The results in Fig. 3a shows that both class 1 and class 9 have the lowest test accuracies of (65.2%) and (76%), respectively. The training accuracies for these two classes are (84.9%) and (87.2%), respectively. This drop in accuracy is due to the fact that they have been widely used in the data across different weather conditions. This made it difficult to identify a distinctive pattern under which these grinds are used. Meanwhile, the remaining grind classes have a small difference between their individual training and testing errors.

B. The second approach: using features clustering

In this approach, the features are clustered using the agglomerative hierarchical clustering method discussed in Section III, resulting in three clusters. The data is resampled using a four-fold cross validation and the ensemble model is trained and tested using one of the three clusters.

Cluster 1 includes snow temperature and humidity as well as air temperature. Training the ensemble model using these features resulted in an averaged training cross validation error of (6.94%) and a testing cross validation error of (12.70%). The individual test accuracies for the ten grinds classes are shown in the confusion matrix in Fig. 3b.

The overall accuracy for the ensemble model trained using this cluster is comparable to the results of the first approach, where all the features are used to train the system. This indicates that the air and snow temperatures along with the snow humidity are the most relevant parameters in determining the grind among the nine snow and weather features investigated in this study. Furthermore, the ensemble model in this case has a higher accuracy in identifying class 1 grind, which is (77.2%) compared to (65.2%) in the first approach.

Cluster 2 includes features that describe the snow conditions, namely: snow hardness, snow grain size, artificial and natural snow types. In this case, training the ensemble model resulted in an averaged training cross validation error of (42.26%) and a testing cross validation error of (44.22%). The test accuracies of the grind classes are given in the confusion matrix shown in Fig. 3c. While the performance of this cluster is poor compared to the first cluster, the ensemble model does not suffer from overfitting as the training and testing errors for the individual grind classes are similar. In addition, this model has the lowest test accuracy for identifying class 1 grind with accuracy of (17.7%).

Cluster 3 includes two features: the air humidity and the precipitation. Training the ensemble model based on these two features resulted in an averaged cross validation error for the training set of (32.24%) and for the testing set of (67.88%). The results show poor performance and overfitting associated with this cluster. This is due to the fact that there are many missing values in the precipitation feature. The testing accuracies for the individual classes are illustrated in the confusion matrix shown in Fig. 3d. It can be noted that, although the testing accuracies for the individual classes are generally low, this ensemble model can recognise class 1 grind with an accuracy of (71.3%). This is because this grind class is associated with high air humidity values and it has few missing values for the precipitation feature.

In summary, comparing the results from the two approaches shows that, generally, the first approach resulted in good training and testing accuracies. Furthermore, it has better performance than the second approach, except when training the ensemble model using cluster 1. The features in cluster 1 can be used to train the ensemble model to predict ski grind classes with a comparable accuracy to the first approach. This shows that air and snow temperatures as well as snow humidity are the key factors in determining the grind classes. In addition, although the ensemble model trained on snow conditions in cluster 2 has a lower overall accuracy, it does not overfit the data. This model has the smallest difference between the training and the testing accuracies. Finally, the features in cluster 3 contain high levels of noise which have increased the chances of overfitting the data. However, training the ensemble model on this cluster can identify grind classes that can be hard to predict based on other features. These classes have distinctive patterns, such as class 1, which is associated with high air humidity.

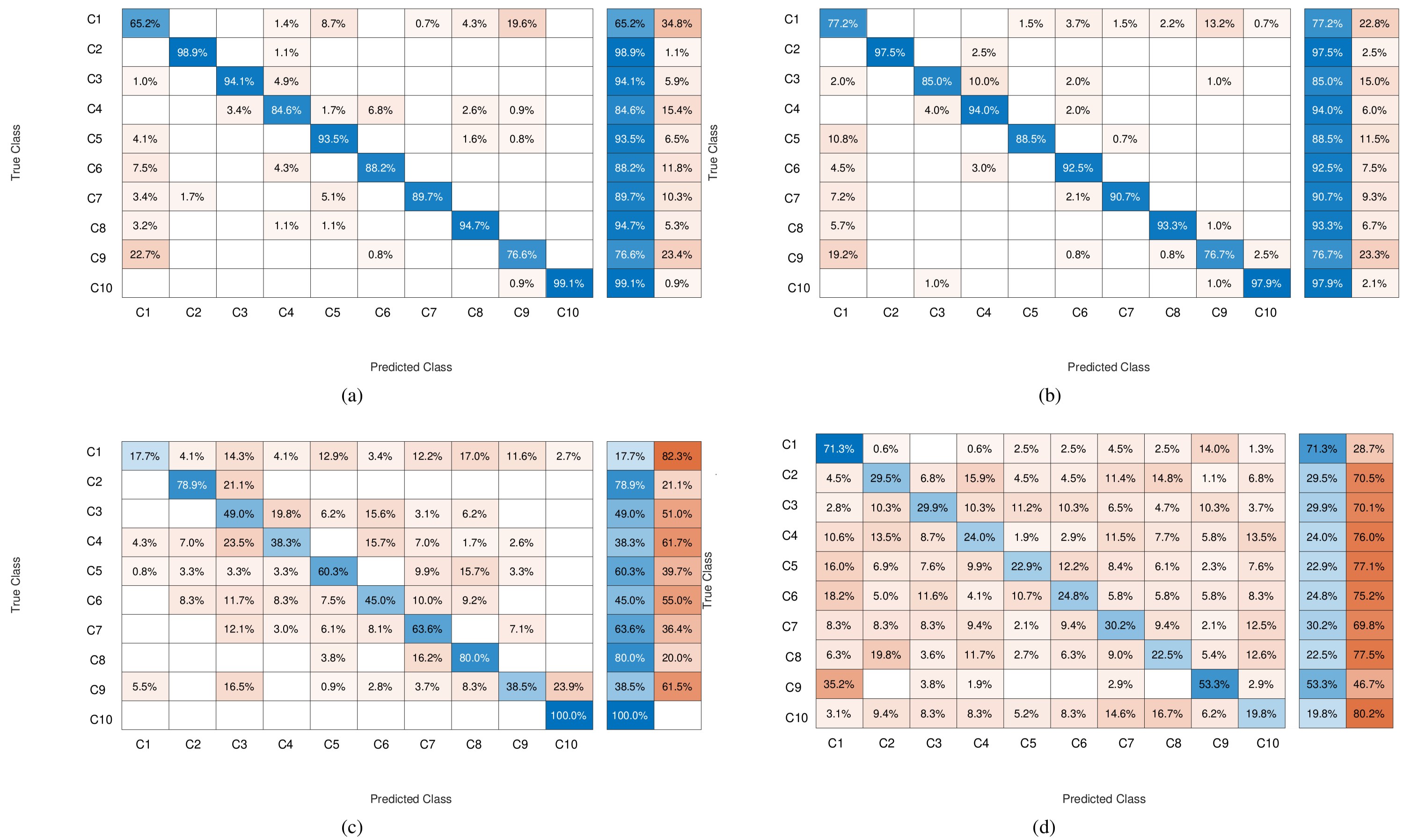

Fig. 3: Confusion matrices for ensemble testing accuracy based on (a) all features (b) cluster 1 (c) cluster 2 and (d) cluster 3, where C1-C10 represent the ten grind classes.

VI. Conclusion and Future Work

This work aimed to develop a better understanding of the friction between snow and skis using machine learning approaches. It had presented an ensemble model that can recommend the best ski grind under different weather conditions. This model can be used as a guiding tool in skiing competitions to aid technicians in choosing skis roughness. Furthermore, this study had analysed the relations among the main nine weather and snow conditions that affect the choice of the ski grind using correlation and clustering approaches. This helped in identifying the features which control the grind selection process. These features were: snow temperature, air temperature, and snow humidity. This is consistent with our understanding that snow friction is related to a thin layer of water acting as a lubricant, resulting from frictional melting [20].

In the future, this work can be expanded to include a wider range of grind classes. Furthermore, though the grind is a property of ski that is chosen carefully to reduce the friction between the snow surface and skis, there are additional more temporary ski treatments that can reduce the friction, for example, the application of different types of wax and the use of rilling tools which cause temporary imprints on the ski base. Future work will focus on the interactions of these treatments with the grind and how they affect the friction between snow and skis.

References

[1] W. Nachbauer, P. Kaps, M. Hasler, and M. Mössner, "Friction between ski and snow," in *The engineering approach to winter sports*. Springer, 2016, pp. 17–32.

[2] P. Carlsson, M. Ainegren, M. Tinnsten, D. Sundström, B. Esping, A. Koptioug, and M. Bäckström, "Cross-country ski," in *The engineering approach to winter sports*. Springer, 2016, pp. 107–151.

[3] E. Sabbioni, S. Melzi, F. Cheli, and F. Braghin, "Bobsleigh and skeleton," in *The Engineering Approach to Winter Sports*. Springer, 2016, pp. 183–276.

[4] W. Nachbauer, P. Schröcksnadel, and B. Lackinger, "Effects of snow and air conditions on ski friction," in *Skiing Trauma and Safety: Tenth Volume*. ASTM International, 1996.

[5] M. Nolte, N. Kister, and M. Maurer, "Assessment of deep convolutional neural networks for road surface classification," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 381–386.

[6] R. Omer and L. Fu, "An automatic image recognition system for winter road surface condition classification," in *13th international IEEE conference on intelligent transportation systems*. IEEE, 2010, pp. 1375–1379.

[7] O. M. H. Rindal, T. M. Seeberg, J. Tjønnås, P. Haugnes, and Ø. Sandbakk, "Automatic classification of subtechniques in classical cross-country skiing using a machine learning algorithm on micro-sensor data," *Sensors*, vol. 18, no. 1, p. 75, 2018.

[8] B. Nemec, T. Petrič, J. Babič, and M. Supej, "Estimation of alpine skier posture using machine learning techniques," *Sensors*, vol. 14, no. 10, pp. 18 898–18 914, 2014.

[9] "Olympiatoppen," https://www.olympiatoppen.no/page943.html , accessed: 2020-09-09.

[10] Y. Qian, E. J. Almazan, and J. H. Elder, "Evaluating features and classifiers for road weather condition analysis," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 4403–4407.

[11] K. Sasirekha and P. Baby, "Agglomerative hierarchical clustering algorithm-a," *International Journal of Scientific and Research Publications*, vol. 83, p. 83, 2013.

[12] R. A. Jacobs, "Methods for combining experts' probability assessments," *Neural computation*, vol. 7, no. 5, pp. 867–888, 1995.

[13] B. Al-Jubouri and B. Gabrys, "Local learning for multilayer, multi-component predictive system," *Procedia Computer Science*, vol. 96, pp. 723–732, 2016.

[14] B. Al Jubouri and B. Gabrys, "Diversity and locality in multi-component, multi-layer predictive systems: A mutual information based approach," in *International Conference on Advanced Data Mining and Applications*. Springer, 2017, pp. 313–325.

[15] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[16] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *European conference on computational learning theory*. Springer, 1995, pp. 23–37.

[17] M. K. Warmuth, J. Liao, and G. Rätsch, "Totally corrective boosting algorithms that maximize the margin," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 1001–1008.

[18] Y. Freund, "A more robust boosting algorithm," *arXiv preprint arXiv:0905.2138*, 2009.

[19] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, no. 8, pp. 832–844, 1998.

[20] S. Colbeck, "The kinetic friction of snow," *Journal of Glaciology*, vol. 34, no. 116, pp. 78–86, 1988.

Acknowledgment

We are thankful to Olympiatoppen for providing the data set used in this study. We are also grateful to Felix Breitschädel who provided expertise that greatly assisted in understanding how the field tests were carried out.