# Intelligent Data-driven Diagnosis of Incipient Inter-turn Short Circuit Fault in Field Winding of Salient Pole Synchronous Generators

Hossein Ehya, *Student Member, IEEE,* Tarjei N. Skreien, and Arne Nysveen, *Senior Member, IEEE*

*Abstract*—This paper examines if machine learning (ML) and signal processing can be used for on-line condition monitoring to reveal inter-turn short circuit fault (ITSC) in the field winding of salient pole synchronous generators (SPSG). This was done by creating several ML classifiers to detect ITSC faults. A data set for ML was built using power spectral density of the air gap magnetic field extracted by fast Fourier transform (FFT), discrete wavelet transform energies, and time series feature extraction based on scalable hypothesis tests (TSFRESH) to extract features from measurements of SPSG operated under several different severities of ITSC fault. Using this data set, a wide range of classifiers were trained to detect the presence of ITSC faults. The classifiers evaluated were logistic regression, K-nearest neighbours, radial basis function support vector machine (SVM), linear SVM, XGBoost decision tree forest, multi-layer perceptron (MLP), and a stacking ensemble classifier including all of the aforementioned. The classifiers were optimised using hyper-parameter grid searches. In addition, some feature selection and reduction algorithms were assessed such as random forest feature selection, TSFRESH feature selection, and principal component analysis. This resulted in a classifier capable of detecting 84.5% of samples containing ITSC fault, with a 92.7% chance that fault detections are correct.

*Index Terms*—Air-gap magnetic field, fault diagnosis, feature extraction, machine learning, signal processing, salient pole synchronous generator.

## I. INTRODUCTION

S ALIENT pole synchronous machines are the machines most commonly used in hydroelectric plants [1], and so are ubiquitous throughout the Norwegian power system. In fact, hydroelectric generation accounted for 95 % of the total electric energy produced in Norway in 2018 [2]. Failure of the synchronous generators that generate the electricity that the Norwegian society is run on can incur not only a great expense in restoring power plants, but also a large cost to society. These machines are under ever-increasing operational demands as intermittent power sources enter the power system. The proper running and maintenance of synchronous machines, and by extension the timely detection and diagnosis of their faults, more important than ever. Hydroelectric generators can suffer failure as a result of undetected incipient faults that induce larger faults. The state-of-the art in on-line fault detection

in salient pole synchronous generators is still lacking in this respect [3], [4].

In the transition from reactive to predictive maintenance, it is vital with accurate estimations of machine states. This involves integrating sensors, signal analysis, and decision-making algorithms. The potential benefits to society are immense, estimated by McKinsey Digital to reach a total potential economic value of 11 trillion USD in 2025 [5], and the power generation sector is no exception. By applying on-line condition monitoring, incipient machine faults can be detected in real-time and faults can be detected before they cause unscheduled stops and further damage to the machine.

The rotor winding inter-turn short-circuit (ITSC) is the failure of insulation between turns in the rotor winding coil so that the number of turns in the coil is effectively reduced [1]. This can be due to overheating causing damage to the insulation, thermal deformation or mechanical stresses [6]. The fault can then propagate to cause the rotor winding to be further short-circuited and eventually a short to ground [6]. Another issue that could arise from the uneven magnetic field is uneven mechanical stresses that further compromise other machine components [6].

The pole-drop test is the most commonly applied off-line test to detect short-circuited turns in the field winding [6]. It is done by applying low voltage AC to terminals of the field winding and measuring the voltage across each pole. A faulty pole will have a lower voltage across it compared to the other poles [6]. The disadvantage of this test is that it requires the machine to be taken off-line. Off-line tests require the shut-down of the machine and are therefore expensive. They also occur while the machine is at a standstill and therefore faults that are induced due to rotational forces can become invisible [6]. To find the faults present during operation, it is necessary to conduct on-line monitoring and tests [6]. On-line condition monitoring for diagnosing rotor winding ITSC is often done using flux probe measurements, where the magnetic field registered by a flux probe placed on a stator tooth in the air gap of the machine is analyzed by comparison to a healthy case [7]. The extracted feature for ITSC fault based on measured air gap magnetic field is done either in steady-state [7] or transient operation of the synchronous machine [8].

Numerous non-invasive approaches exist to diagnose ITSC fault in SPSG based on stray flux analysis [9], harmonics of stator current and voltage [10], and unbalanced circulating current in the stator winding [11]. In [12], a sensorless method

Hossein Ehya, Tarjei N. Skreien, and Arne Nysveen are with Department of Electrical Power Engineering, Norwegian University of Science and Technology, Trondheim, 7030, Norway. (corresponding author phone: +47 47743322; e-mail: hossein.ehya@ntnu.no).
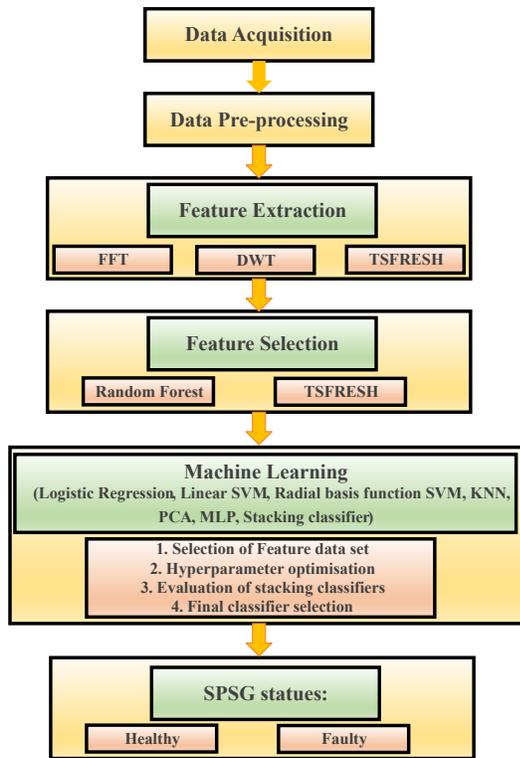
Fig. 1. Procedure of health statues determination of SPSG based on intelligent data-driven approach.

TABLE I
SPECIFICATION OF 100 $kVA$, 50 $Hz$, SYNCHRONOUS GENERATOR

| Quantity | Values | Quantity | Values |
|---|---|---|---|
| No. of slots | 114 | No. of damper bars/pole | 7 |
| Winding connection | Wye | Number of poles | 14 |
| No. of stator turns | 8 | No. of rotor turns / pole | 35 |
| Nominal speed | 428 rpm | Power factor | 0.90 |
| Nominal voltage | 400 V | Nominal current | 144.3 A |
| Nominal exc. current | 103 A | No-load exc. current | 53.2 A |
| Nominal exc. voltage | 20 V | No-load exc. Voltage | 10.5 V |

and classifier robustness is increased by integration of various base learners in order to form an ensemble learner [28], [29].

This paper applies ensemble stacking classifiers in combination with a sparse sensor application of a single air-gap magnetic flux sensor to detect ITSC faults. This combination of several ML algorithms into one improves predictive performance, while the single sensor is minimally invasive. Previous applications have used solitary ML models, whereas in this paper it is shown that a superior result can be achieved by *combining* ML models. Furthermore, generating feature-rich data sets using automatic feature generation algorithms makes the procedure nearly sensor agnostic. This system is applied on data that is pre-processed to resemble samples likely to be found in industry to avoid over-confident performance assertions.

In order to investigate which machine learning models perform the best, and if a single air gap magnetic field sensor is sufficient for reliable ITSC fault diagnosis or not, a fault classification system as shown in 1 has been created in this paper that includes:

1) Automatic sample processing and segmentation from longer sample series
2) A feature extraction process capable of processing and organising an arbitrary number of samples
3) A feature selection process that employs several feature selection methods
4) A process to assess the usefulness of feature selection, select the best machine learning model among several, and assess the performance of the final model
5) A final ensemble classifier to detect ITSC faults

based on measurements of the induced voltage in the screw located in the stator core was proposed. Although the induced voltage in the screw is mirrored the air gap magnetic field of the synchronous generator, the method is not sensitive to a low degree of ITSC. The magnetic field of the rotor shaft, shaft voltage, and its current are also proposed in [13], where it does not have adequate sensitivity to less severe ITSC faults in synchronous generator rotor poles.

Numerous signal processing tools based on a frequency domain or joint time-frequency domain were applied to the aforementioned signals and different features were extracted to identify the ITSC fault. However, interpretation of the data requires an expert in the field. Therefore, data-driven methods may exclude prior knowledge that is suitable for fault detection. Artificial intelligence has become a useful technique which may be employed in data-driven fault detection of electrical machines. Both supervised and unsupervised machine learning approaches have demonstrated their effectiveness in fault diagnosis [14]. Unsupervised methods are trained on unlabeled data, and are frequently used in fault classification. K-means and self-organizing neural networks such as ART networks in combination with wavelet are used for fault detection purposes [15]–[17]. Support vector machine (SVM) [18]–[20], K-nearest neighbor (KNN) and artificial neural networks (ANN) [17], [21]–[23], fuzzy logic network [24], principal component analysis [25], [26], convolutional neural network [20] and XGBoost [27] are widely used as a supervised machine learning classifiers for fault diagnosis of electric machines. Although mentioned methods showed their ability in classification of fault in electrical machine, accuracy

## II. LABORATORY TEST

### A. Experimental Set-up

The data set is composed of two concurrent Hall-effect sensor readings taken of a salient pole synchronous generator running at synchronous speed in no-load and full-load with several different ITSC-fault severities induced. The machine, the sensors attached and the measurements are described as below:

1) A 100 $kVA$, 400 $V$, synchronous generator with 14 salient poles constructed to resemble generators commonly situated in Norwegian hydroelectric power plants. It is shown in Fig. 2. Its nameplate value and some defining features of its topology are given in Table I.
2) The generator was driven by a 90 $kW$, 400 $V$ induction motor with four poles and rated speed of 1482
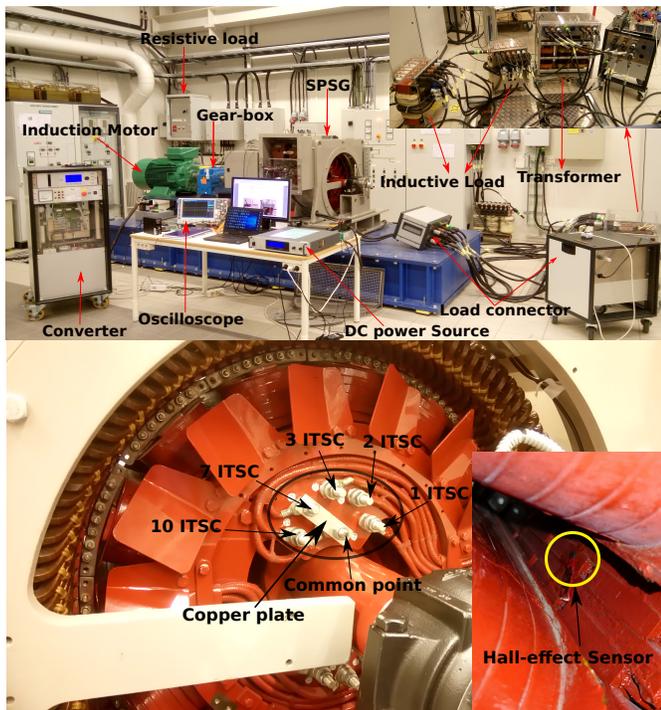
Fig. 2. The experimental test rig of a 100 $kVA$ synchronous generator (top). A copper shunt utilized to short circuit the field winding of SPSG. The location of installed hall-effect sensor on a stator tooth in yellow circle (bottom).
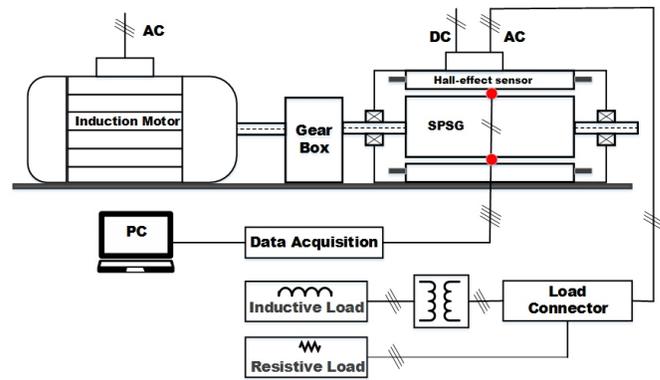


Fig. 3. The connection diagram of experimental set-up.

TABLE II
THE SPECIFICATION OF GENERATOR FROM NO-LOAD TO FULL-LOAD IN
FOUR DIFFERENT CASES IN THE HEALTHY OPERATION OF SPSG

| Characteristics | No-load | Load 1 | Load 2 | Load 3 |
|---|---|---|---|---|
| Output Power | - | 30 $kVA$ | 40 $kVA$ | 65 $kVA$ |
| Power Factor | - | 1.0 | 0.79 | 0.93 |
| Exc. Current | 56 $A$ | 60 $A$ | 77 $A$ | 84 $A$ |
| Exc. Voltage | 10.5 $V$ | 11.1 $V$ | 13.9 $V$ | 14.7 $V$ |
| $I_R$ | 0 $A$ | 43 $A$ | 43 $A$ | 84 $A$ |
| $I_L$ | 0 $A$ | 0 $A$ | 36 $A$ | 33 $A$ |

rpm supplied by a three-phase converter. The speed of the induction motor during all tests was set so that the frequency of the generator's electrical output was 50.004 $Hz$.

3) A gearbox was used to connect the shaft of the induction motor to the synchronous generator.

4) A programmable converter was employed to control the operation of the induction motor. The converter is supplied by an external rectifier connected to the grid.

5) A 20 $kW$ (LAB-HP/E2020) DC power source was utilized to magnetize the field winding.

6) Two Hall-effect sensors (AST244) were placed into the air gap and glued onto stator teeth at diametrically opposing ends of the stator as shown in Fig. 2. The dimensions of the sensors were $(3.0 \times 5.0 \times 0.8)mm$ with a flux density to a voltage ratio of 2.54 $T/V$. The constant DC current supply is used to feed 4.75 $mA$ into the Hall-effect sensors. The data sheet specifies that the sensor should be supplied by a 2 $mA$ DC current source. However, due to considerable electromagnetic interference, the magnitude of the current power supply was increased to 4 $mA$ to increase the signal-to-noise ratio.

7) A high-resolution oscilloscope (16-bit Tektronix MSO 3014), with a sampling frequency of 10, and 50 $kHz$, was used for data acquisition.

8) A water-cooled resistor comprised of two parallel sets of resistors, where the total resistance can be controlled and adjusted in steps by contactors and relays in a separate control panel. The per-phase resistance could be varied

from a maximum of 160 $\Omega$ to a minimum of about 2.78 $\Omega$. At the maximum load setting, the dissipated power of the resistors amounts to about 57 $kW$.

9) Two inductive loads, in which each phase are connected in series, are connected to the generator by a three-phase transformer. The approximate value of the inductance in each phase based on the turn ratio of the transformer is equal to 22 $mH$.

10) A copper plate was used to make an ITSC fault on one of rotor field winding by short circuiting 1, 2, 3, 7, or/ 10 turns as shown in Fig. 2

### B. Test Procedure

Fig. 3 presents a connection diagram of the experimental test rig. The Procedure of experimental tests is as follows: tests were performed in healthy and faulty cases in no-load, fully resistive, and resistive-inductive load according to the table. II. The SPSG which was coupled to an induction motor with a help of a gearbox is accelerated until it reaches its nominal synchronous speed. The magnetizing current is increased until the stator voltage reaches its nominal value. The magnetizing current is increased by increasing the load to maintain the stator voltage in its nominal value. ITSC fault is conducted at standstill by removing a certain desired number of turns from the rotor field winding with a help of a copper plate. As shown in Fig. 2 there exists a common tap on a rotor field winding connected to a bolt located at the rotor which is called a common point. There are 5 taps on the rotor field winding that is connected to the bolts that enable to apply ITSC fault by removing 1, 2, 3, 7, and 10 turns. For instance, by connecting a common point and tap (7 ITSC) as shown in Fig. 2 by using the copper plate, 7 turns are removed. In total
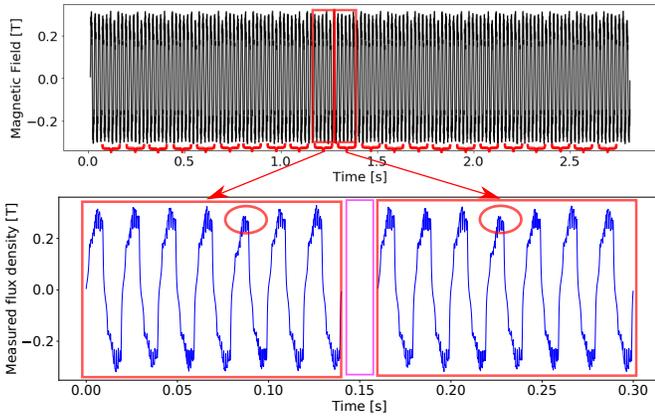
Fig. 4. Two consecutive RSS that each of them represents one data set cut from the same OSS. They are each of 7 electrical periods, with a 1 electrical period between the two. Note the smaller negative peak occurring in periods 4 and 3 occurring of the first and second RSS respectively. The one period shift between each RSS makes the fault indication appear one position earlier.

48 experiments were conducted, each of which sampled with two sensors and two sampling frequencies. The temperature effects on tests were examined by comparing the results in the cold and warm operation of SPSG. Analysis shows that temperature does not have an impact on the acquired signal.

## III. METHOD AND RESULTS

### A. Data Pre-processing

The data was processed to appear similar to something one would sample in a production environment. In a production deployment of the fault detection system, the measurement series would need to be windowed with the classification run on a sliding window of the last electrical periods to be able to detect faults in near real-time. Since incipient faults are not critical, a long window length of several mechanical periods is possible. The minimum viable window length is 1 mechanical period, as this is the window length necessary to ensure that any fault will pass the sensors. An excessively long window length is prohibitive since it will add little new information and slow down feature extraction. However, the window length should be long enough to remediate end effects in signal processing tools that suffer from them. End effects can be alleviated by analysing a concatenated series if the signal is assumed to be periodic. Since the machine has 7 pole pairs, 7 electrical periods will capture 1 complete mechanical period. The reduced sample series (RSS) extracted from original sample series (OSS) are cut at rising zero-crossing to have integer electrical periods in each RSS as shown in Fig. 4. Each RSS represents a data set, indicating that, the total number of RSS used for ML purposes in this paper by performing 48 experiments (8 healthy cases and 40 faulty cases) under different load conditions and fault severity is approximately equal to 2500.

### B. Feature Extraction

To generate features, signal processing methods can be used in concert with discipline knowledge. From a frequency

#### TABLE III
THE THREE DATA SETS TAKEN INTO MACHINE LEARNING.

| Set | Selection method | Num. features |
|-----|------------------|---------------|
| A | None | 417 |
| B | Random forest | 81 |
| C | TSFRESH | 301 |

spectrum generated by a signal processing method, one would select the frequencies of the signal that are most informative and generate some features from that. This could be the energy spectrum of a certain decomposition level in discrete wavelet transform, the intensity of some side-band frequencies relative to a harmonic frequency, or any other property of the signal or its transforms.

Raw time series are very sensitive to small perturbations and thus not suited to be used directly as tabular training data. Features are therefore extracted from each RSS that are then used as a basis for feature selection and, finally, as training data. The feature extraction methods used were fast Fourier transform, discrete wavelet transform energies and TSFRESH feature extraction. In total 475 distinct features were extracted.

*1) Fast Fourier Transform:* The frequency content of each RSS was extracted by FFT. FFTs of healthy and faulty signals showed that the faulty signal had a marked increase in harmonic frequency components at intervals of $f_m = \frac{50}{7} \ Hz$, the mechanical frequency of the generator, outside of the odd multiples of fundamental frequency compared to the healthy case. The frequency components of integer multiples of $f_m$ up to 500 $Hz$ were extracted as features, see 1.

$$f_{k,extracted} = k \cdot f_m = k \cdot \frac{2f_{sync}}{p}, k = 0, 1, 3, \ldots \quad (1)$$

*2) DWT wavelet energies:* A 12-level-decomposition, Haar wavelet DWT was taken of each RSS and instantaneous, Teager, hierarchical, and relative wavelet energies were computed for each decomposition level. An issue with DWT is its end effects, which are worsened substantially in each decomposition level since the length of the data series that is transformed is effectively halved in each decomposition level with the Haar wavelet. The adverse effects are diminished as the length of the data series increases since the portions affected by end effects are proportionally smaller. Therefore, each RSS was concatenated to 4 times its length before the DWT was taken. This exploits the assumption that the generator behaviour is stationary.

*3) TSFRESH:* An algorithm to extract features from time series, called FeatuRe Extraction based on Scalable Hypothesis tests (FRESH) is proposed in [30]. Its intent is to automate time series feature extraction while implementing feature selection. The FRESH algorithm was integrated into a algorithmic feature generation package, called Time Series FeatuRe Extraction based on Scalable Hypothesis tests (TSFRESH) [31]. TSFRESH is able to generate a total of 794 time series features, using 63 time series characterisation methods as well as applying feature selection methods. A comprehensive feature extraction was done using TSFRESH. TSFRESH's

FFT features were not included because TSFRESH did not offer the ability to select frequencies of interest.

## IV. FEATURE SELECTION

Two feature selection methods, random forest feature selection and TSFRESH, were applied to the feature data set. Before any feature selection was undertaken, a hold-out data set was extracted from it to prevent any target leakage.

### A. Random forest feature selection

The random forest feature selection was done using a forest of 1000 decision tree estimators, which were trained on the training set using Gini impurity as the splitting criterion. During training, every feature is assigned an importance based on its impurity. All features of greater than mean importance were selected, the remainder discarded. This resulted in a feature reduction from 417 to 81 features.

### B. Time series feature extraction based on TSFRESH

Using the feature extraction module included in TSFRESH, a subset of features deemed relevant was extracted. TSFRESH was configured to assume dependent features. False discovery rates in the interval 0.001, 0.01, 0.05, and 0.1 were tried, this resulted in a similar amount of features. The false discovery rate settled upon was 0.05, the rate used in [32]. This resulted in a feature reduction from 417 to 301 features.

### C. Selected Data Sets

The three versions of the feature data set, hereby termed feature data sets A, B and C, are summarised in Table III. By comparing the performance of classifiers trained upon the different collections of features, some insight into which features are most useful for classifying the fault can be gleaned and which feature selection algorithms are most useful with this data. In a final version of the fault detection system, this knowledge could be used to selectively compute only the most useful features.

## V. MACHINE LEARNING

The following section details the development of a classifier intended to detect the presence of ITSCs using the data sets previously created. This is done in four phases:

1) Selection of the feature data set
2) Hyperparameter optimisation of single machine learning models
3) Evaluation of stacking classifiers
4) Final classifier selection and evaluation on hold-out data set

The first objective, selection of the feature data set, was accomplished by evaluating the results of training a host of different classifiers on each data set. The classifiers chosen were:

1) Logistic Regression with and without PCA
2) KNN with and without PCA
3) Radial basis function SVM with and without PCA

4) Linear SVM with and without PCA
5) XGBoost
6) MLP
7) Stacking classifier

By implementing logistic regression, SVM and KNN with and without a PCA, the effectiveness of PCA in this application can be gauged as well. PCA was not combined with XGBoost because PCA reduces the interpretability of the model, a key strength of decision trees. The PCA was identically executed in all four applications.

*1) Evaluation Metrics:* There are several ways to evaluate the performance of classifiers, and they give differing results. Perhaps the simplest method is to count the number of correct classifications and divide by the total number of samples. This is what is called the *accuracy* of the classifier, shown in 2. It says something about the performance of the classifier, but has trouble with unbalanced data sets. Given an unbalanced electric machine measurement data set containing 99% of samples of healthy machines and 1% of samples of faulty machines, a classifier that always classifies a sample to be healthy would have a 99% accuracy. This is obviously a poor classifier as it would never correctly classify a single faulty machine. This is addressed by including other measurements that also emphasise the misclassified samples. A popular metric that does this is the *F-score*. It works by combining *sensitivity, and precision.*

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad (2)$$

A useful tool to talk about these measures is the confusion matrix for a binary classifier that classifies samples as belonging to the class, true, or not belonging to the class, false. The confusion matrix contains the number of samples that are: correctly classified as belonging to the class, true positive (TP); incorrectly classified as belonging to the class, false positive (FP); incorrectly classified as not belonging to the class, false negative (FN); and correctly classified as not belonging to the class, true negative (TN).

Sensitivity, shown in 3, is a measure of how well the model picks up on the class, essentially the probability that the class is detected.

$$sensitivity = \frac{TP}{TP + FN} \qquad (3)$$

Specificity, shown in 4, gives an impression of the model's capacity to correctly classify false samples.

$$specificity = \frac{TN}{TN + FP} \qquad (4)$$

Precision, shown in 5, is the ratio of true positives divided by the total number of samples classified as true. A high precision gives confidence that the classifier has made a correct prediction when it returns true.

$$precision = \frac{TP}{TP + FP} \qquad (5)$$

Each of these has pit-falls when faced with unbalanced data sets and classifiers that classify all samples as either
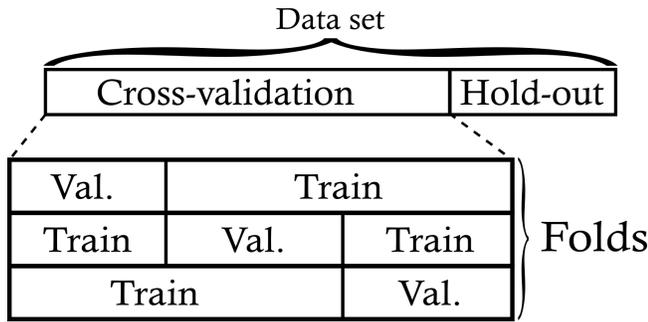
Fig. 5. Three-fold cross-validation. Each fold is composed of a training and validation set.

TABLE IV
A SUMMARY OF THE RESULTS OF CLASSIFIERS TRAINED ON DATA SETS.

| Data set | Classifier | Sensitivity | Precision |
|---|---|---|---|
| A | Logistic Regression | 0.8853 | 0.7722 |
| | Logistic Reg. with PCA | 0.8622 | 0.8131 |
| | KNN | 0.8269 | 0.6747 |
| | KNN with PCA | 0.8201 | 0.6775 |
| | SVM (rbf) | 0.8492 | 0.7050 |
| | SVM (rbf) with PCA | 0.8538 | 0.6312 |
| | SVM (linear) | 0.8859 | 0.7612 |
| | SVM (linear) with PCA | 0.8576 | 0.8176 |
| | XGBoost | 0.8518 | 0.7766 |
| | Multi-layer Perceptron | 0.8833 | 0.7390 |
| | Stack | 0.8652 | 0.8191 |
| | Average classifier score | 0.8583 | 0.7443 |
| B | Logistic Regression | 0.8675 | 0.7772 |
| | Logistic Reg. with PCA | 0.8140 | 0.7394 |
| | KNN | 0.8074 | 0.7237 |
| | KNN with PCA | 0.8392 | 0.7207 |
| | SVM (rbf) | 0.8117 | 0.7029 |
| | SVM (rbf) with PCA | 0.8149 | 0.6453 |
| | SVM (linear) | 0.8790 | 0.7925 |
| | SVM (linear) with PCA | 0.7878 | 0.7392 |
| | XGBoost | 0.8407 | 0.7193 |
| | Multi-layer Perceptron | 0.8702 | 0.7322 |
| | Stack | 0.8712 | 0.7981 |
| | Average classifier score | 0.8367 | 0.7355 |
| C | Logistic Regression | 0.8966 | 0.7998 |
| | Logistic Reg. with PCA | 0.8663 | 0.8082 |
| | KNN | 0.8282 | 0.6743 |
| | KNN with PCA | 0.8222 | 0.6756 |
| | SVM (rbf) | 0.8531 | 0.7226 |
| | SVM (rbf) with PCA | 0.8492 | 0.6327 |
| | SVM (linear) | 0.8972 | 0.8106 |
| | SVM (linear) with PCA | 0.8615 | 0.8162 |
| | XGBoost | 0.8313 | 0.7816 |
| | Multi-layer Perceptron | 0.8859 | 0.7643 |
| | Stack | 0.8714 | 0.8485 |
| | Average classifier score | 0.8603 | 0.7577 |

true or false. To balance the possible pitfalls, the F-score is especially good for unbalanced classes. The F-score is defined as the harmonic mean of precision and sensitivity, it weighs the reliability of a classification together with its chance of detecting the class [33]. The F1-score is shown in 6.

$$F_1 - score = 2 \cdot \frac{precision \cdot sensitivity}{precision + sensitivity} \qquad (6)$$

*2) Cross-validation:* Since data sets are not entirely uniform, the results of the train/test procedure are affected by the way the data is split. One split may by chance give very good test results, while another does the opposite. This could result in selecting a model that generalises poorly even though it performs well on the test set. To counter this, k-fold cross-validation can be used [34]. k-fold cross-validation takes in a data set and splits it into k folds. Each fold is composed of a training set and a validation set. For each fold, the model is trained on the training set and its performance measured on the validation set. The models performance is then the average performance across all the folds, and the performance is more likely to reflect the true performance of the model on unseen data.

Since what is of interest when testing a new model is its performance on *new and unseen data*, a part of the data set should be set aside to be used only to assess the performance of the model. This is known as a *hold-out data-set* (as shown in Fig. 5) that includes 20% of the data set. The data sets were previously, during the feature selection process, split into a hold-out test data set and a remainder data set. Since the results of a single train/test cycle can be very dependant upon the split of the samples, the classifiers were evaluated by their average performance across a 5-fold CV. This produces 5 folds of CV-train and CV-validation sets drawn from the remainder data set of the initial split. The folds are identical across all classifiers and feature data sets.

*3) Standardisation:* Logistic Regression, KNN and SVM are sensitive to the variance of the samples, this is addressed by applying standardisation. Each cross-validation split was standardised to zero-mean and unity variance. The mean and variance of every feature was calculated from the CV-train set. Both CV-test and CV-validation sets were standardised using the CV-train means and variances.

*4) Results:* This procedure was repeated for every classifier on every feature data set and performance metrics were

gathered. The results are presented in Table IV. This method of model fitting was used for every classifier evaluation at later stages of the classifier development.

### A. Feature selection and reduction performance

It appears that the choice of data set does not greatly affect the performance of the classifiers, and the variance of the results is large. However, feature data set C, the TSFRESH feature selection data set, slightly outperforms the rest on every averaged metric. Data set C is thus preferred, and will be utilised from this point onward.

As for feature reduction, i.e. application of PCA, every classifier suffered a drop in performance in nearly every metric when PCA was applied. Of special note is that radial basis function SVM with PCA had an ROC AUC consistently lower than 0.5, which indicates that it performed worse than chance. Due to this, PCA was abandoned. It might still have been justified on grounds of reducing training and prediction time if there were more features or an extremely large number of samples, but no such considerations were necessary.

### B. Hyperparameter optimisation and selection

Since a classifier's performance is heavily dependent upon its hyperparameters, all the candidate classifiers were opti-

TABLE V
HYPERPARAMETER SEARCH GRIDS FOR LOGISTIC REGRESSION, KNN, SVM, AND XGBOOST CLASSIFIERS.

| Classifier | Hyperparameter | Values | Final Value | Description |
|---|---|---|---|---|
| Log. Reg. | C | $10^k$, $k = -10, -9.5, ..., 10$ | $10^{8.5}$ | Inverse of regularisation strength |
| | penalty | "l1", "l2" | 12 | Penalisation norm |
| KNN | n_neighbors | $1, 3, 5, ..., 351$ | 1 | Number of nearest neighbours |
| SVM | C | $10^k$, $k = -1, 0, 1, 2, 3$ | 10 | Inverse of regularisation strength |
| | gamma | $10^k$, $k = 0, -1, -2, -3, -4$ | 1 | Inverse of regularisation strength |
| | kernel | "rbf", "linear" | linear | Kernel type |
| XGBoost | learning_rate | $0.01, 0.2, 0.3, 0.5$ | 0.5 | Learning rate |
| | n_estimators | $100, 400, 700, 1000$ | 100 | Number of trees in ensemble |
| | max_depth | $3, 10, 15, 25$ | 3 | Maximum tree depth |
| | col_sample_bytree | $0.8, 1$ | 0.8 | Per tree column subsampling ratio |
| | subsample | $0.6, 0.8, 1$ | 1 | Sample subsampling ratio |
| | reg_alpha | $0.7, 1, 1.3$ | 1.3 | Lasso regularisation term on weights |
| | reg_lambda | $0, 0.5, 1$ | 0 | Ridge regularisation term on weights |
| MLP | activation | 'identity', 'logistic', 'tanh', 'relu' | identity | The activation function |
| | batch_size | $200, 133, 66, 32$ | 200 | Size of minibatches |
| | max_iter | $200, 500, 1000, 1200$ | 200 | The maximum number of epochs |
| | hidden_layer_sizes | (50,25,3), (100,50,7), (200,100,14), (300,150,21) | (50, 25, 3) | Size and number of hidden layers |

TABLE VI
THE ACCURACY, SENSITIVITY, PRECISION AND F1-SCORE OF THE BEST
MODELS FOUND IN THE HYPER-PARAMETER GRID SEARCH.

| Classifier | Accuracy | Sensitivity | Precision | F1-score |
|---|---|---|---|---|
| Logistic Regression | 0.7986 | 0.8740 | 0.8376 | 0.8506 |
| KNN | 0.6395 | 0.8350 | 0.6990 | 0.7501 |
| SVM | 0.7940 | 0.8854 | 0.8247 | 0.8501 |
| XGBoost | 0.7438 | 0.8576 | 0.7846 | 0.8142 |
| MLP | 0.7958 | 0.9022 | 0.8170 | 0.8542 |

TABLE VII
THE RESULTS FROM THE STACKING CLASSIFIER COMPARISON.

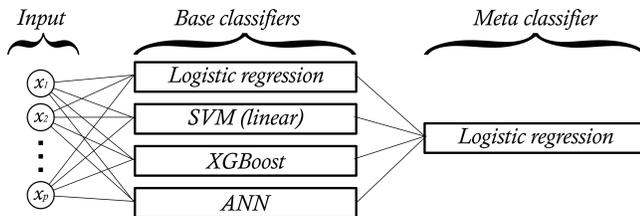| Meta-classifier | Accuracy | Sensitivity | Precision | F1-score |
|---|---|---|---|---|
| Logistic Regression | 0.7840 | 0.8701 | 0.8260 | 0.8432 |
| Multi-layer Perceptron | 0.7479 | 0.8057 | 0.8276 | 0.8107 |
| Gradient boosting forest | 0.7663 | 0.8268 | 0.8304 | 0.8255 |
| Random Forest | 0.7704 | 0.8216 | 0.8388 | 0.8265 |



Fig. 6. A stacking classifier with Logistic Regression as its meta-classifier.

mised before selecting among them. The optimisation procedure was a 5-fold cross-validating grid search. In this procedure, a hyperparameter grid is defined that contains a range of values for each of the hyperparameters to be optimised. The grid search algorithm then executes a cross-validation of the classifier for every possible combination of these hyperparameters. The mean cross-validation performance is calculated for each hyperparameter combination, and the hyperparameter combination that yields the best performance on the chosen performance metric is selected. The performance metric used is F1-score because it combines sensitivity and precision.

The hyperparameter sets with the greatest mean performance across 5-fold cross-validation for each classifier are presented as final value in Table V. Table VI shows the scores of these classifiers across several metrics. Of the optimised classifiers, the XGBoost and KNN models are outperformed by the others. KNN's accuracy was 64.0% in an imbalanced data set of 65.9% majority class. This performance is worse than that of a dummy classifier that classifies randomly or always classifying samples as the majority class. Furthermore, KNN is entirely non-generalising with a $k = 1$, indicating that the algorithm is not well suited for this problem.

*C. Ensemble learners*

Ensemble learners are learners that combine several weak learners that may have poor performance to create a stronger learner with better performance. There are a few methods of accomplishing this, mainly *bagging, boosting, and stacking*. Stacking is to train a meta-learner, a model that is trained to interpret the outputs of several other models to make a prediction based on the predictions of the other learners. The learners that provide predictions to the meta-learner are termed base-learners. It usually outperforms the base-learners it is trained upon. Each of the base-learners are first fitted to the training set, and their predictions upon the training set are used as the training set for the meta-learner. The base-learners can be any machine learning model that returns predictions. This provides a benefit in that by including different models as base-learners, the weaknesses of one model can be remedied by another.

Since a stacking classifier improved the performance during the feature data set selection, the same approach is made again using the optimised classifiers. Four stacking classifiers were made with different meta-classifiers, Logistic Regression, MLP, gradient boosting forest, and a random forest classifier. The gradient boosting forest classifier was chosen over XGBoost as a meta-classifier due to greater compatibility with Sci-kit Learn's stacking framework. Since XGBoost is

TABLE VIII
THE RESULTS OF THE BEST OF THE SINGLE AND STACKING CLASSIFIERS
ON THE HOLD-OUT DATA SAMPLES.

| Classifier | Accuracy | Sensitivity | Precision | F1-score |
|---|---|---|---|---|
| Logistic Regression | 0.7569 | 0.6961 | 0.9435 | 0.8011 |
| Logistic Reg. stack | 0.8448 | 0.8456 | 0.9274 | 0.8846 |

also a variant of gradient boosting forest, it should return similar results at the expense of computing power. The stacks all include the optimised Logistic Regression, SVM, MLP, and XGBoost classifiers as base classifiers (Fig. 6). KNN was again excluded due to its poor performance and slow prediction time. Results are shown in Table VII. Of the stacking classifiers, the Logistic Regression stacking classifier outperformed the others by a large margin.

Comparing the performance of the best stacking classifier with that of the best non-ensemble classifier, a somewhat surprising result surfaces. The logistic regression classifier alone on average slightly outperforms the stacking classifier of which it is a part of across the cross-validation folds.

An advantage of stacking classifiers is that they often generalise better than single classifiers, and they usually outperform their base classifiers. However, the hyperparameters of the meta-classifier have not been optimised on the training set as is the case with the simple logistic regression classifier. To gauge their performance on unseen samples, both are trained on the entire training set and tested on the hold-out data set. The results are presented in Table VIII.

On the hold-out set, the stacking classifier outperforms the simple logistic regression classifier. The stacking classifier could likely be further improved by running a grid-search for the optimal hyper parameters of the logistic regression meta-classifier.

## VI. CONCLUSION

This paper has investigated how signal processing and machine learning tools can be used to detect inter-turn short-circuits in rotor field windings. This was done in three stages, data pre-processing, feature extraction and selection, and classifier development as described below:

1) Signal partitioning is used to achieve a sufficient number of data sets to train the intelligent system.
2) The features extracted were power spectral density of integer multiples of the generator's mechanical frequency extracted by FFT, DWT wavelet energies, and the entire TSFRESH feature extraction suite. The most useful features were the wavelet energy features and some of the TSFRESH features.
3) Linear machine learning models were best suited for fault detection on this data set, especially the logistic regression and linear SVM classifiers. The best classifier was an ensemble stacking classifier with logistic regression as the meta-classifier taking inputs from logistic regression, XGBoost, linear SVM, and MLP classifiers as base-classifiers.

The results indicate that ITSC fault classification using machine learning on air-gap magnetic field measurements from a single sensor can yield good results. The logistic regression stacking classifier had an accuracy of 0.8448, a sensitivity of 0.8456, and a precision of 0.9274. This means that the classifier correctly classified 84.48% of all the samples in the hold-out data set, and 84.56% of the faulty samples present were correctly classified as such. Of the samples that were classified as faulty, 92.74% were correctly classified. Since a large portion of faults go undetected, this fault detection system should therefore not be relied upon as the only detection system. However, if the system alerts of a fault, it would warrant investigation since it is likely to be correct.

A general trend during optimisation was that linear machine learning models performed well and that the performance of non-ensemble classifiers *increased* as the complexity *decreased*. The worst performance was exhibited by the K-nearest neighbours classifier, performing worse than random chance.

Future work in this research includes:

1) Using a combination of various sources of signals such as vibration and stray magnetic field to achieve higher accuracy and sensitivity in a classifier.
2) ITSC fault severity assessment using some of the same methodology mentioned in this paper.
3) Application of ensemble stacking classifier in different kinds of faults in synchronous generators such as eccentricity fault, and broken damper fault.

## REFERENCES

[1] D. P. Kothari and I. J. Nagrath, *Electric Machines*. Tata McGraw-Hill Education, 2004, google-Books-ID: axGw7r3SOEMC.

[2] "08583: Elektrisitetsbalanse (MWh) 2010m01 - 2019m09." [Online]. Available: http://www.ssb.no/statbank/table/08583/

[3] H. Ehya, I. Sadeghi, and J. Faiz, "Online condition monitoring of large synchronous generator under eccentricity fault," in *2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2017, pp. 19–24.

[4] I. Sadeghi, H. Ehya, J. Faiz, and A. A. S. Akmal, "Online condition monitoring of large synchronous generator under short circuit fault — a review," in *2018 IEEE International Conference on Industrial Technology (ICIT)*, 2018, pp. 1843–1848.

[5] J. Manyika, J. Woetzel, R. Dobbs, M. Chui, P. Bisson, J. Bughin, and D. Aharon. Unlocking the potential of the internet of things | McKinsey. Library Catalog: www.mckinsey.com. [Online]. Available: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world

[6] J. Yun, S. Park, C. Yang, Y. Park, S. B. Lee, M. Šašić, and G. C. Stone, "Comprehensive Monitoring of Field Winding Short Circuits for Salient Pole Synchronous Motors," *IEEE Transactions on Energy Conversion*, vol. 34, no. 3, pp. 1686–1694, Sep. 2019.

[7] J. A. Antonino-Daviu, M. Riera-Guasp, J. Pons-Llinares, J. Roger-Folch, R. B. Pérez, and C. Charlton-Pérez, "Toward Condition Monitoring of Damper Windings in Synchronous Motors via EMD Analysis," *IEEE Transactions on Energy Conversion*, vol. 27, no. 2, pp. 432–439, Jun. 2012.

[8] Y. Park, S. B. Lee, J. Yun, M. Sasic, and G. C. Stone, "Air gap flux-based detection and classification of damper bar and field winding faults in salient pole synchronous motors," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 3506–3515, 2020.

[9] M. Cuevas, R. Romary, J. Lecointe, F. Morganti, and T. Jacq, "Noninvasive detection of winding short-circuit faults in salient pole synchronous machine with squirrel-cage damper," *IEEE Transactions on Industry Applications*, vol. 54, no. 6, pp. 5988–5997, 2018.

[10] M. Valavi, K. G. Jørstad, and A. Nysveen, "Electromagnetic analysis and electrical signature-based detection of rotor inter-turn faults in salient-pole synchronous machine," *IEEE Transactions on Magnetics*, vol. 54, no. 9, pp. 1–9, 2018.

[11] L. Hao, Y. Sun, A. Qiu, and X. Wang, "Steady-state calculation and online monitoring of interturn short circuit of field windings in synchronous machines," *IEEE Transactions on Energy Conversion*, vol. 27, no. 1, pp. 128–138, 2012.

[12] W. Yucai, M. Qianqian, and C. Bochong, "Fault diagnosis of rotor winding inter-turn short circuit for sensorless synchronous generator through screw," *IET Electric Power Applications*, vol. 11, no. 8, pp. 1475–1482, 2017.

[13] J. S. Hsu and J. Stein, "Shaft signals of salient-pole synchronous machines for eccentricity and shorted-field-coil detections," *IEEE Transactions on Energy Conversion*, vol. 9, no. 3, pp. 572–578, 1994.

[14] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2226–2238, 2013.

[15] B. Chen, X. Wang, S. Yang, and C. McGreavy, "Application of wavelets and neural networks to diagnostic system development, 1, feature extraction," *Computers Chemical Engineering*, vol. 23, no. 7, pp. 899 – 906, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0098135499002586

[16] S. M. Cruz and A. M. Cardoso, "Multiple reference frames theory: A new method for the diagnosis of stator faults in three-phase induction motors," *IEEE Transactions on Energy Conversion*, vol. 20, no. 3, pp. 611–619, 2005.

[17] A. Glowacz and Z. Glowacz, "Diagnosis of the three-phase induction motor using thermal imaging," *Infrared physics & technology*, vol. 81, pp. 7–16, 2017.

[18] B. M. Ebrahimi, M. Javan Roshtkhari, J. Faiz, and S. V. Khatami, "Advanced eccentricity fault recognition in permanent magnet synchronous motors using stator current signature analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 4, pp. 2041–2052, 2014.

[19] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560 – 2574, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327007000027

[20] S. Munikoti, L. Das, B. Natarajan, and B. Srinivasan, "Data-driven approaches for diagnosis of incipient faults in dc motors," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5299–5308, 2019.

[21] P. Janik and T. Lobos, "Automated classification of power-quality disturbances using svm and rbf networks," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1663–1669, 2006.

[22] A. Glowacz, "Fault diagnostics of acoustic signals of loaded synchronous motor using smofs-25-expanded and selected classifiers," *Tehnički vjesnik*, vol. 23, no. 5, pp. 1365–1372, 2016.

[23] A. Glowacz, W. Glowacz, Z. Glowacz, and J. Kozik, "Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals," *Measurement*, vol. 113, pp. 1 – 9, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0263224117305432

[24] F. Zidani, D. Diallo, M. E. H. Benbouzid, and R. Nait-Said, "A fuzzy-based approach for the diagnosis of fault modes in a voltage-fed pwm inverter induction motor drive," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 2, pp. 586–593, 2008.

[25] A. Stief, J. R. Ottewill, J. Baranowski, and M. Orkisz, "A pca and two-stage bayesian sensor fusion approach for diagnosing electrical and mechanical faults in induction motors," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9510–9520, 2019.

[26] W. Zhao and L. Wang, "Multiple-kernel mrvm with lbfo algorithm for fault diagnosis of broken rotor bar in induction motor," *IEEE Access*, vol. 7, pp. 182 173–182 184, 2019.

[27] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and xgboost," *IEEE Access*, vol. 6, pp. 21 020–21 031, 2018.

[28] Z. Xu, C. Hu, F. Yang, S. Kuo, C. Goh, A. Gupta, and S. Nadarajan, "Data-driven inter-turn short circuit fault detection in induction machines," *IEEE Access*, vol. 5, pp. 25 055–25 068, 2017.

[29] M. Seera, C. P. Lim, S. Nahavandi, and C. K. Loo, "Condition monitoring of induction motors: A review and an application of an ensemble of hybrid intelligent models," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4891 – 4903, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417414000918

[30] M. Christ, A. W. Kempa-Liehr, and M. Feindt, "Distributed and parallel time series feature extraction for industrial big data applications," *arXiv:1610.07717 [cs]*, May 2017, arXiv: 1610.07717. [Online]. Available: http://arxiv.org/abs/1610.07717

[31] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)," *Neurocomputing*, vol. 307, pp. 72–77, Sep. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231218304843

[32] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," vol. 57, no. 1, pp. 289–300, _eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1995.tb02031.x. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x

[33] D. D. Lewis, R. E. Schapire, J. P. Callan, and R. Papka, "Training algorithms for linear text classifiers," in *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 298–306.

[34] R. R. Picard and R. D. Cook, "Cross-validation of regression models," vol. 79, no. 387, pp. 575–583, publisher: Taylor & Francis Group.

**Hossein Ehya** (S'19) received the M.Sc. degree in electrical engineering from the Department of Electrical and Computer Engineering, University of Tehran, Tehran, Iran, in 2013. From 2013 to 2018, he worked as an electrical machine design engineer in electrical machine companies.

He is currently working toward the Ph.D. degree in electrical engineering at the Norwegian University of Science and Technology (NTNU), Trondheim, Norway. In 2020, he was awarded with the ICEM Jorma Luomi Award in Gothenburg, Sweden. His research interests include the design and condition monitoring of electrical machines, signal processing, pattern recognition, and machine learning.

**Tarjei Nesbø Skreien** is a master student graduate from The Norwegian University of Technology and Science in Trondheim, Norway. His research interests include synchronous machine fault detection, machine learning and sampling noise mitigation.

**Arne Nysveen** (M'98–SM'06) received his Dr. ing. degree (PhD) and his MSc in electric power engineering from the Norwegian Institute of Technology (now NTNU), Trondheim, Norway, in 1994 and 1988, respectively.

From 1995 to 2002, he was a Senior Scientist with ABB Corporate Research, Oslo, Norway. Since 2002, he has been a Professor at the Norwegian University of Science and Technology (NTNU), Trondheim. He is currently manager for the research on Turbine and Generator Technologies at the Norwegian Research Center for Hydropower Technology (HydroCen). His current research activities are on design, modelling and monitoring of hydroelectric generators.