

# Störschallunterdrückung bei Luftschallanalysen in industriellen Fertigungsstrecken

Johannes Nowak<sup>1</sup>, Sascha Grollmisch<sup>1</sup>, Estefanía Cano<sup>1</sup>, Hanna Lukashevich<sup>1</sup>  
und Judith Liebetrau<sup>1</sup>

<sup>1</sup>Fraunhofer-Institut für Digitale Medientechnologie IDMT, 98693 Ilmenau, Deutschland, Email: noa@idmt.fraunhofer.de

## Einleitung

Dieser Artikel beschreibt Untersuchungen zur automatisierten akustischen Qualitätskontrolle (ACM – Acoustic Condition Monitoring) von Elektromotoren in industriellen Fertigungsstrecken. Um Probleme in der Fertigung bereits frühzeitig zu identifizieren, nutzen gängige ACM-Verfahren Vibrationsaufnehmer (sogenannte Körperschallmikrofone) für die Analyse. Besonders bei Prüflingen mit mechanischen, beweglichen Bauteilen, wie Elektromotoren, kündigen sich Fertigungsungenauigkeiten oder -fehler bereits frühzeitig im Klangbild an. Die akustische Analyse des Klangbildes erlaubt die Klassifikation, ob ein Prüfling *in Ordnung* oder *nicht in Ordnung* ist, und bietet somit die Möglichkeit rechtzeitig entsprechende Gegenmaßnahmen einzuleiten. Aus diesem Grund ist ACM ein wichtiger Bestandteil bei der automatisierten Qualitätskontrolle im Bereich Industrie 4.0.

In praktischen ACM-Anwendungen haben Körperschallmikrofone jedoch Nachteile: erstens müssen diese in einem zusätzlichen Verarbeitungsschritt an den Prüfling angebracht werden, was zu einer kostenintensiven Erhöhung der Taktfrequenz in der Fertigung führt; zweitens erlauben Körperschallmikrofone nur die Analyse eines begrenzten spektralen und räumlichen Bereichs; und drittens nehmen diese Sensoren auch Körperstörschall aus der Umgebung auf und verfälschen somit die Klassifikationsergebnisse. Um diese Probleme in den Griff zu bekommen, ist die Industrie an ACM-Verfahren interessiert, die auf Luftschallanalysen beruhen. Während die Luftschallanalysen unter idealen Laborbedingungen bereits gute Ergebnisse liefern, so sind diese aufgrund der starken Störgeräusche in einer realen Fertigungsstrecke bzw. Werkshalle jedoch bisher nicht praxistauglich.

Im vorliegenden Artikel werden realitätsnahe Experimente durchgeführt und verschiedene Verfahren getestet und bewertet, um ein ACM-System gegen den Einfluss von Störschall zu härten. Es werden drei Ansätze verfolgt: im ersten Ansatz wird das Stör-Nutzsignalverhältnis (SNR – Signal to Noise Ratio) bereits bei der Aufnahme durch den Einsatz von Richtmikrofonen optimiert; im zweiten Ansatz werden die maschinellen Lernverfahren (ML) für die Zustandsklassifikation gezielt gegen Störschall trainiert; im dritten Ansatz werden sogenannte Quellentrennverfahren (QT) aus dem Bereich des Music Information Retrieval (MIR) eingesetzt, um den Störschalleinfluss zu reduzieren und somit das Nutzsignal hervorzuheben. Die Güte der getesteten Ansätze werden anhand der erreichten Klassifikationsgenauigkeit analysiert und bewertet.

## Grundlagen

Dieser Artikel fokussiert auf ACM-Systeme, die den charakteristischen Klang eines Elektromotors automatisch mithilfe von ML-Verfahren klassifizieren [13, 5, 2]. Dafür werden die Grundlagen im Folgenden kurz vorgestellt, wie auch für die Anwendung von QT-Verfahren in industriellen Umgebungen. Durch Fortschritte im Bereich des ML und der erfolgreichen Anwendung in einer Vielzahl an Gebieten, verspricht die Kombination von ACM-Systemen mit Methoden des ML großes Potential [9, 8, 12]. Jedoch wird die Genauigkeit der Erkennung stark durch störende Signalanteile beeinflusst. Robustere ML-Verfahren wie tiefe neuronale Netze (DNN – Deep Neural Networks) können diesen negativen Einfluss verringern [14]. Darüber hinaus werden in aktuellen ACM-Ansätzen Methoden der QT angewandt, um das Nutzsignal von Störanteilen zu säubern [4, 11] und somit die Klassifikationsgenauigkeit zu erhöhen.

## Maschinelles Lernen

Maschinelles Lernen kann grundlegend in drei Ansätze unterteilt werden: „supervised“, „semi-supervised“ und „unsupervised“ [3]. Während für supervised-Methoden alle Daten vollständig annotiert vorliegen, arbeiten semi-supervised-Verfahren mit nur teilweise annotierten Daten. Liegen keinerlei Datenannotationen vor, so werden unsupervised-Verfahren angewandt, welche die Daten automatisch in Bereiche (Cluster) aufteilen. Im vorliegenden Artikel werden supervised-Verfahren angewandt. Hierbei existieren unterschiedliche Algorithmen, die je nach Anwendungsfall diverse Vor- und Nachteile mit sich bringen. In den letzten Jahren konnten sich DNNs in vielen Bereichen gegenüber bisherigen Ansätzen durchsetzen, weshalb diese auch für den vorliegenden Artikel verwendet wurden. Obwohl künstliche neuronale Netze bereits in den 40er und 50er Jahren entwickelt wurden, fanden diese erst in den letzten Jahren in immer mehr Forschungsgebieten, wie z.B. der Bildklassifikation, Anwendung [6, 7]. DNNs sind von der Signalverarbeitung im menschlichen Gehirn inspiriert und bestehen aus mehreren Schichten von Neuronen, die durch gewichtete Verbindungen zu den vorherigen Schichten aktiviert werden. Die erste Schicht erhält die Merkmale (z.B. ein Spektrogramm) als Eingang, während die Ausgabe der letzten Schicht den Ergebnissen entspricht. Die Gewichte in den Schichten werden innerhalb der Trainingsphase anhand von Trainingsmerkmalen erlernt. Die erreichbare Klassifikationsgüte hängt hierbei maßgeblich von der Parameterwahl und der Netzwerkarchitektur ab [6].

## Akustische Quellentrennung

Obwohl die Verfahren der QT ursprünglich für den Bereich des MIR entwickelt wurden, liegt der generelle Anwendungsfall im Trennen von akustischen Quellen aus einem gemischten Audiosignal. Besonders Ansätze, die keine Annahme bezüglich harmonischer oder rhythmischer Eigenschaften treffen, bieten sich für die Anwendung auf industriellen Audiodaten an, um die Nutzsignalanteile aus dem Mischsignal zu filtern. Aktuelle Untersuchungen zeigten die generelle Eignung dieser Verfahren für industrielle Geräusche [4].

In den Experimenten, die in diesem Artikel beschrieben werden, wird der Algorithmus Kernel Additive Modeling for Interference Reduction (KAMIR) genutzt, der sich speziell für die Anwendung in Mehrkanalaufnahmen eignet. KAMIR basiert auf dem Kernel Additive Model (KAM) [1] und versucht das Übersprechen der Kanäle zu reduzieren. Dabei setzt KAMIR voraus, dass jede Schallquelle einem eigenen Kanal zugewiesen ist. In diesem Fall werden die Hauptmikrofone auf den Prüfling ausgerichtet, während die restlichen Mikrofone den Störschall erfassen sollen. Durch den Einsatz eines verallgemeinerten Wiener-Filters werden die Quellen iterativ getrennt [11].

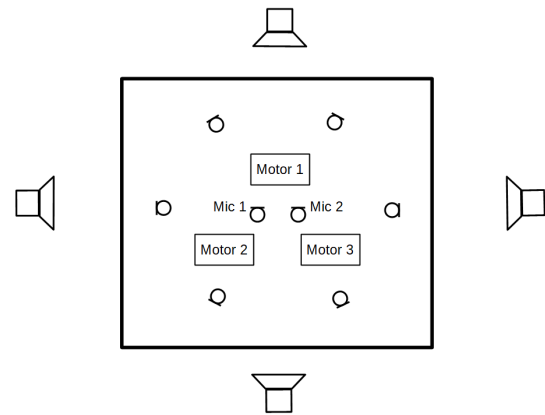
## Experiment

Im Rahmen des Experimentes wird der Einfluss von Störschall auf die Klassifikationsgenauigkeit von Elektromotoren untersucht und dabei verschiedene Möglichkeiten evaluiert, die Stör-Nutzsignaltrennung zu optimieren und somit die Klassifikationsgenauigkeit zu verbessern. In den folgenden Abschnitten werden die Testumgebung und -konditionen beschrieben, die Klassifikationsergebnisse für die unterschiedlichen Konditionen präsentiert und die Ergebnisse diskutiert.

## Testumgebung

Für die Untersuchungen wurde eine industrielle Fertigungszelle mit Aluminiumgestell (auch Boschprofil genannt) und Plexiglasscheiben nachgebildet. In der Mitte der Zelle wurden die Prüflinge positioniert: drei baugleiche Elektromotoren, die jeweils unterschiedliche Klangcharakteristika aufweisen: *in Ordnung*, *nicht in Ordnung* und *unter Last*. Die klanglichen Unterschiede sind hörbar und nur einer der Motoren ist gleichzeitig aktiv. Innerhalb der Fertigungszelle wurden acht Mikrofone angebracht, die mit Schwanenhälsen fest positioniert wurden. Von den acht Mikrofonen fungierten zwei als Hauptmikrofone (Mic 1 und Mic 2) zur Aufnahme der Elektromotoren (Nutzschall) und sechs als Störschallmikrofone, um den Umgebungslärm aufzeichnen. Die Störschallmikrofone sind explizit für die QT zur Optimierung des Stör-Nutzsignalverhältnisses vorgesehen. In den Experimenten wurden Mikrofone mit Kugel- und Nierencharakteristik verwendet, um den Einfluss der Mikrofonrichtcharakteristik zu bewerten. Dafür wurden Sennheiser MKH 800 Mikrofone eingesetzt, bei denen die

Richtcharakteristik entsprechend eingestellt wurde. Der Störschall wird mit fünf Lautsprechern simuliert, von denen einer oberhalb der Zelle positioniert war, um elektrischen Umgebungslärm zu simulieren. Die übrigen vier Lautsprecher wurden an jeder Seite der Fertigungszelle platziert. Der experimentelle Aufbau ist schematisch in Abbildung 1 dargestellt.



**Abbildung 1:** Schematische Darstellung des experimentellen Aufbaus, mit der Fertigungszelle, den drei Motoren und den acht Mikrofonen, und den Lautsprechern für die Simulation von Störschall.

Als Störschallsignal dient realer Umgebungslärm, welcher in einer Werkshalle mit einem omnidirektionalen Messmikrofon (Earthworks M30) aufgezeichnet wurde. Insgesamt stehen knapp über 30 Minuten Audiomaterial zur Verfügung. Das Störschallsignal wurde über alle Lautsprecher gleichzeitig wiedergegeben, wobei jeder Lautsprecherkanal zufällig verzögert wurde, um Phasenartefakte zu verhindern. Der Umgebungslärm wird in den Untersuchungen mit fünf unterschiedlichen Lautstärken wiedergegeben: *kein*, *schwach*, *mittel*, *stark* und *extrem*. Jeder Störschallpegel wurde mit jedem Motor jeweils 10 Minuten aufgezeichnet, was einer Gesamtdauer des Datensatzes von ungefähr 150 Minuten pro Mikrofon entspricht. Die Abtastrate betrug 48 kHz.

## Klassifikation

In dem durchgeführten Experiment handelt es sich um ein 3-Klassenproblem, da nur einer der drei Motoren gleichzeitig aktiv sein kann. Entsprechend wurde jede Aufnahme hinsichtlich Störschallpegel und aktivem Motor annotiert. Für die Klassifikation wird ein DNN verwendet. Als Eingabemerkmale dienen die spektralen Frames, die durch eine STFT mit einer Fensterlänge von 2048 Samples und einer Sprungweite von 512 Samples ermittelt wurden. Da sich die Klänge der Motoren zeitlich stabil verhalten und sich besonders in ihrem Frequenzmuster unterscheiden, wurden die Magnituden durch einen laufenden Mittelwert geglättet. Eine Fenstergröße von 15 Frames (160 ms) erzielte in Voruntersuchungen die besten Ergebnisse (die Glättung entspricht einer Reduzierung des Einflusses von kurzzeitig auftretendem Störschall).

Das DNN umfasst zwei versteckte Schichten mit 512 und 64 Neuronen (Feed-Forward). Zusätzliche Schichten bzw. Neuronen brachten in den Untersuchungen keinen Zugewinn an Klassifikationsgenauigkeit. Die Initialisierung der Gewichte erfolgte zufällig. Zusätzlich wurden Dropout-Schichten nach jeder versteckten Schicht eingefügt, um Übertraining zu verhindern. Eine Dropout-Wahrscheinlichkeit von 0.5 erzielte dabei die besten Resultate. Als Aktivierungsfunktion wurde ReLU (Rectified Linear Unit) [10] in den versteckten Schichten verwendet. Als Optimierungsalgorithmus kam Adam mit einer Lernrate von 0.001 zum Einsatz, der die Trainingsphasen beschleunigen sollte. Als Kostenfunktion wurde Categorical Cross Entropy verwendet. Die Implementierung erfolgte in Python3 unter Nutzung von Keras <sup>1</sup> mit Tensorflow<sup>2</sup>.

Die Aufnahmen wurden bei jeder Untersuchung in Trainings-, Validierungs- und Testdaten aufgeteilt. Die Trainingsdaten dienten zum Erlernen der Gewichte des DNN; die Auswahl des besten Modells erfolgte anhand der ungesehenen Validierungsdaten; und der finale Test des Modells auf den ebenfalls ungesehenen Testdaten. Als Validierungsdaten wurden zufällig 20 % der Aufnahmen aus den Trainingsdaten genutzt. Die Jede Untersuchung wurde mit 30-facher Kreuzvalidierung durchgeführt, wobei in jedem Schritt die Verteilung von Trainings- und Validierungsdaten neu bestimmt wurde. Als Gütemaß wurde die Genauigkeit pro Frame bestimmt, d.h. die Klasse mit der höchsten Genauigkeit wird mit der annotierte Klasse verglichen. Die angegebenen Ergebnisse entsprechen dem Mittelwert der Einzelergebnisse.

Die folgende Tabelle 1 gibt einen Überblick über die erreichten Klassifikationsergebnisse ohne QT. Das DNN wurde hierfür nur mit den Aufnahmen ohne Störschall (*kein*) trainiert:

**Tabelle 1:** Klassifikationsergebnisse des 3-Klassenproblems für unterschiedliche Mikrofone.

Störschall	Niere		Kugel	
	Mic 1	Mic 2	Mic 1	Mic 2
Kein	99,99 %	99,99 %	99,99 %	99,99 %
Schwach	99,99 %	99,99 %	99,99 %	99,99 %
Mittel	99,99 %	99,99 %	99,99 %	99,99 %
Stark	97 %	98 %	92 %	96 %
Extrem	81 %	96 %	68 %	63 %

Die Ergebnisse zeigen, dass bei *keinem*, *geringem* oder *mittleren* Störschall bereits sehr hohe Klassifikationsgenauigkeiten von 99.99 % erreicht werden können. Für diesen Fall hat die Richtcharakteristik des Mikrofons und seine Position nur einen geringen Einfluss auf die Klassifikation. Bei *starkem* Störschallpegel nimmt die Klassifikationsgenauigkeit deutlich ab und der Einfluss der Richtcharakteristik steigt: Während mit Nierenmikrofonen noch Genauigkeiten von 97 % und 98 % erreicht werden, fällt die Genauigkeit für Mic 1 auf 92 % und die für Mic 2 auf 96 %. Der Effekt wird noch deutlicher,

wenn *extremer* Störschall auf die Mikrofone trifft. Hier fällt die Klassifikationsgenauigkeit bei Nierenmikrofonen bis auf 81 % oder gar 63 % bei Kugelmikrofonen. Diese Ergebnisse zeigen, dass die höhere Richtwirkung der Nieren bereits eine bessere Stör-Nutzsignaloptimierung erlaubt und dass die Positionierung der Mikrofone eine Rolle spielt. Generell lässt sich festzustellen, dass die Erkennung von Motor *unter Last* für alle Untersuchungen die größten Probleme bereitet. Dies kann entweder an der räumlichen Anordnung der Motoren zueinander liegen, oder an der größeren Ähnlichkeit des Klanges zu den anderen beiden Klassen.

Für die weiteren Analysen wird aufgrund der schlechteren Ergebnisse – und dem damit größeren Optimierungspotential – lediglich Mic 1 mit zeitlicher Glättung herangezogen und zwei weitere Ansätze untersucht, um die Klassifikationsgenauigkeit zu erhöhen: Im ersten Ansatz wird das DNN zusätzlich mit *geringem* Störschall trainiert. In der Evaluation zeigt sich, dass das DNN dadurch gegen *starken* und *extremen* Störschall gehärtet werden kann. Dies erklärt sich durch die größere Varianz der Trainingsdaten, wodurch das gelernte Modell stärker generalisiert. In allen Fällen konnte damit die Klassifikationsgenauigkeit auf 99.99 % gesteigert werden. Im zweiten Ansatz wird der KAMIR-Algorithmus angewandt, um das Stör-Nutzsignalverhältnis zu optimieren. Hierbei wurde die Anzahl der genutzten Mikrofone für die Trennung schrittweise erhöht und die Anzahl virtueller Schallquellen auf 2 gesetzt (Mic 1 plus Störschallmikrofone). Die Hinzunahme weiterer virtueller Schallquellen führt zu einer Verschlechterung, da auch wichtige Frequenzanteile der Elektromotoren gefiltert wurden. Getestet wurde auf den Aufnahmen mit extremen Störschall und trainiert ohne Störschall. Die Referenzgenauigkeit ohne Quellentrennung liegt bei 81% für Nierenmikrofone und 67% für Kugelmikrofone für Mic 1. Tabelle 2 zeigt die Ergebnisse in Abhängigkeit der verwendeten Anzahl an Störschallmikrofonen.

**Tabelle 2:** Klassifikationsergebnisse unter Anwendung der QT.

Störschallmikrofone	Niere	Kugel
2	84 %	67 %
4	84 %	67 %
6	85 %	67 %

Die Ergebnisse zeigen, dass bereits mit 2 zusätzlichen Störschallmikrofonen unter Anwendung des KAMIR-Algorithmus die Klassifikationsgenauigkeit um 3% verbessert werden kann, ohne dass das erlernte Modell geändert werden muss, oder Wissen über den Prüfling für die QT notwendig ist. Bei Hinzunahme weiterer Störschallmikrofone steigt der Rechenaufwand stark, ohne jedoch die Genauigkeit zu erhöhen. Bei Kugelmikrofonen hat dieser Ansatz, wie anzunehmen, keinen Einfluss auf die Genauigkeit, da durch die fehlende Richtcharakteristik Nutz- und Störschall annähernd gleich stark in allen Kanälen vorhanden sind.

<sup>1</sup><https://keras.io>

<sup>2</sup><https://www.tensorflow.org>

## Zusammenfassung

Dieser Artikel beschreibt ein Experiment zur Klassifikation von drei Elektromotoren basierend auf Luftschallanalysen unter Berücksichtigung von Störschall. Die Motoren waren im Nachbau einer Fertigungszelle positioniert. Der Störschall, basierend auf Aufnahmen einer realen Werkshalle, wurde mit fünf Lautsprechern wiedergegeben, welche um die Fertigungszelle positioniert waren, wobei der Störschall mit fünf unterschiedlichen Lautstärken ausgespielt wurde. Für die Klassifikation der drei Motoren wurde ein DNN verwendet und darüber hinaus auch Quellentrennverfahren angewandt, um in *lauten* oder *extremen* Störschallsituationen den Einfluss des Störschalls zu unterdrücken und somit die Klassifikationsgenauigkeit zu erhöhen. Die Ergebnisse zeigen, dass bei *keinem*, *geringem* oder *mittleren* Störschall bereits sehr hohe Genauigkeiten von über 99% erzielt werden können. Wie zu erwarten, nimmt die Klassifikationsgenauigkeit aber bei *starken* und *extremen* Störschallpegeln ab (97% bzw. 81%), besonders wenn nur die störschallfreien Daten für das Training verwendet werden. Werden zusätzlich noch die Daten mit *wenig* Störschall ins Training miteinbezogen, so steigt die Erkennungsgenauigkeit auch bei *extremen* Störschall auf über 99%. Damit wird deutlich, dass das DNN bereits durch geeignete Trainingsdaten gegen den Einfluss von Störschall gehärtet werden kann. In einem weiteren Experiment wurden Quellentrennverfahren, konkret das KAMIR-Verfahren, für die Störschallunterdrückung eingesetzt. Auch hier zeigen die Ergebnisse, dass damit die Klassifikationsgenauigkeit um 3–7% gesteigert werden kann, ohne eine Veränderung des DNN zu benötigen. Auch wurde gezeigt, dass die Verwendung von Richtmikrofonen bei der Aufnahme schon den SNR verbessern und somit die Klassifikationsgenauigkeit des DNNs erhöhen.

Während die beschriebenen Untersuchungsergebnisse nahe legen, dass die Luftschallprüfungen in industriellen Fertigungsstrecken auch mit lautem Störschallpegel möglich ist, bleiben noch einige Fragen unbeantwortet: Kann das untersuchte Verhalten auch in realen Produktionsstätten nachgebildet werden? Wie verhält sich die Erkennung und QT bei zeitlich variierenden Geräuschen bzw. bei komplexeren Klassifikationsaufgaben? Existieren effektivere QT-Verfahren für industrielle Geräusche?

## Danksagung

Die Autoren möchten der Firma GÖPEL Electronics GmbH für ihre Unterstützung und die Bereitstellung der Fertigungszelle danken, insbesondere den Personen Frank Pauli, Jens Jaeger und Oliver Frischmuth. Weiterer Dank gilt Paul Fröhmer für die Durchführung der Messungen und der entsprechenden Datenvorbereitung.

## Literatur

- [1] A. Liutkus, D. Fitzgerald, Z. Rafii, B. Pardo, and L. Daudet. Kernel additive models for source separation. *IEEE Trans. Signal Process.*, 62(16):4298–4310, 2014.
- [2] A. Purarjomandlangrudi and G. Nourbakhsh. Acoustic emission condition monitoring: An application for wind turbine fault detection. *Int. J. Res. Eng. Technol. (IJRET)*, 2(5):907–918, 2013.
- [3] C. M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer-Verlag New York Inc, 1 edition, 2006.
- [4] E. Cano, J. Nowak, and S. Grollmisch. Exploring sound source separation for acoustic condition monitoring in industrial scenarios. In *2017 24th European Signal Processing Conference (EUSIPCO)*, Kos Island, Greece, 2017.
- [5] F. Elasha, M. Greaves, D. Mba, and A. Addali. Application of acoustic emission in diagnostic of bearing faults within a helicopter gearbox. In *Proc. 4th Int. Conf. Through-life Eng. Services (CIRP)*, volume 38, pages 30–36, ChangChun, China, 2015.
- [6] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.
- [7] D. Hassabis. Alphago: using machine learning to master the ancient game of go. <https://blog.google/topics/machine-learning/alphago-machine-learning-game-go/>, January 2016.
- [8] J. H. Zhou, C. K. Pang, Z. W. Zhong, F. L. Lewis, and J.F. Jabuz. Tool wear monitoring using acoustic emissions by dominant-feature identification. *IEEE Trans. Instrum. Meas.*, 60(2):547–559, 2011.
- [9] N. F. Ince, C.-S. Cao, M. Kaveh, A. Tewfik, and J.F. Jabuz. A machine learning approach for locating acoustic emission. *EURASIP J. Advances Signal Process.*, 2010(15), 2010.
- [10] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.
- [11] T. Prätzlich, R. M. Bittner, A. Liutkus, and M. Müller. Kernel additive modeling for interference reduction in multi-channel music recordings. In *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pages 584–588, Brisbane, Australia, 2015.
- [12] W. L. Woon, A. El-Hag, and M. Harbaji. Machine learning techniques for robust classification of partial discharges in oil–paper insulation systems. *IET Sci. Meas. & Technol.*, 10(3):221–227, 2016.
- [13] X. Liu, X. Wu, and C. Liu. A comparison of acoustic emission and vibration on bearing fault detection. In *Proc. IEEE Int. Conf. Transportation Mech. Elect. Eng. (TMEE2011)*, pages 922–926, ChangChun, China, 2011.
- [14] Y. Lecun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.