

RESEARCH PROPOSAL

TITLE

An Investigation into the Applicability of Tiny
Machine Learning (TinyML) for Welsh
Manufacturing SMEs

STUDENT NUMBER

2126963

COURSE NUMBER

ACFA7004

DATE

10/10/22

Executive Summary

In Wales, the biggest challenges regarding SMEs adopting Industry 4.0 technology are the cultural/resistance to change and the lack of knowledge of the applications in which AI can be used.

A new promising research field in tiny machine learning (TinyML) is paving way as a method to reduce the barrier of entry for SMEs to adopt AI to improve their productivity, quality and processes. TinyML focuses on combining machine learning and embedded systems (IoT devices) to provide applications which are executed on small, real-time, low-power and low-cost embedded devices.

This research proposes an investigation into the applicability of adopting TinyML for Welsh manufacturing SMEs. There are two parts to this research, firstly reviewing where TinyML can be implemented in the Welsh manufacturing sector and secondly, conducting a representable manufacturing use case using TinyML through commercial platforms and low-cost technology. The outcome of this research will dictate whether TinyML is ready to be a viable technology entry for Welsh SMEs to benefit their manufacturing processes.

Preliminary literature review is conducted to understand the type of manufacturing challenges SMEs are generally facing, the benefits and challenges of TinyML, the applications where TinyML can be used, the hardware and software which creates a TinyML system, deployment methodologies and hardware benchmark tests.

The method for this research will be mixed, conducting a survey to distribute to Welsh manufacturing SMEs to capture their challenges and an experiment at the AMRC shopfloor to design and implement an TinyML system which performs anomaly detection on an industrial conveyor. This system would then be subjected to further experiments to test the robustness and reliability of the system in an industrial environment.

The results of the survey will be used to map TinyML applications (through extended literature review) to indicate businesses where the technology can be used to reduce their challenges. The results of the implemented system will be assessed against key KPIs and the methodology will be written up in a report to disseminate this knowledge across SMEs. The research project spans across 8 months and the total cost of the hardware and software needed to conduct the project is £511.

Table of Contents

Introduction.....	4
Introduction to Industry 4.0	4
Definition of Artificial Intelligence.....	4
Significance of AI in Wales	5
AI Adoption Challenges for Welsh Manufacturing SMEs	6
Introduction into TinyML	6
Research Problem	9
Literature Review.....	11
Manufacturing Challenges for SMEs.....	11
Applications of TinyML.....	13
Benefits and Challenges with TinyML	17
Hardware, Software and Models for TinyML Applications.....	21
TinyML Deployment Methodologies	24
TinyML Benchmarking	25
Research Aims and Objectives	26
Research Scope.....	27
Proposal	27
Methods	27
Survey	28
Experiment.....	29
Results	33
Timeline	35
Costs.....	36
Proposal Impact.....	37
References.....	38
Appendix A	42

Introduction

Introduction to Industry 4.0

The rise of Industry 4.0 has been revolutionising the way companies manufacture, improve and distribute their products (IBM, 2022) through key technologies which fuse digital, physical and biological worlds together. By definition, Industry 4.0 is a name for the current trend of automation and data exchange in manufacturing technologies. It includes cyber-physical systems, the Internet of things, cloud computing and cognitive computing (AMRC, 2022). In Wales, the Welsh Government (WG) are keen in supporting manufacturing companies, in particular, small and medium sized enterprises (SMEs) to adopt Industry 4.0 technologies with the aim of creating a resilient manufacturing economy and enabling those businesses to be better equipped and flexible in dealing with high and low volumes and multiple product variants (Welsh Government, 2021).

Definition of Artificial Intelligence

This research proposal delves into the field of Artificial Intelligence (AI). AI is defined as the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision making, and translation between languages (Tyagi, 2020). There are six branches of AI (shown in **Appendix A**) including machine learning (ML) and neural networks (NN), which will be the focus of this proposal. Machine Learning is the science of getting computers to learn without being explicitly programmed (Ng, 2022) and neural networks is a subset of ML, inspired by neurons in a human brain, which allows the computer to learn from large amounts of data (IBM Cloud Education, 2020). The relationship between these three definitions is illustrated in **Figure 1** (Shrivastav, 2020).

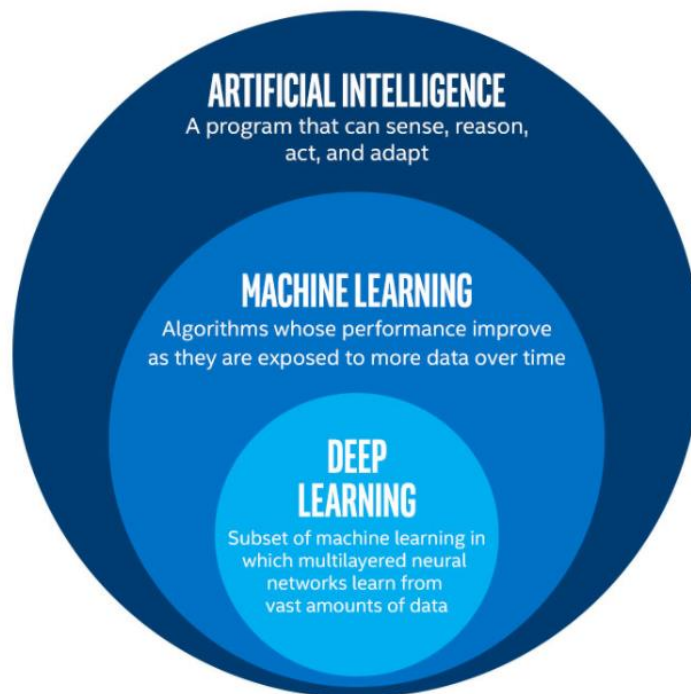


Figure 1: Relationship between AI, ML and DL (NN)

Significance of AI in Wales

In a 2019 WG review on Industry 4.0 technology document, the WG highlighted the importance of AI in delivering economic transformation. The review recommends the establishment of an AI Institute for the Future Economy with an estimated budget requirement of £60 million (Brown, 2019). This signifies the effort which is being placed to roll out AI to Welsh businesses. The focus is on Welsh SMEs because 71% of all Welsh private sector employment is generated by SMEs, including 63% of business turnover in Wales (FSB Wales, 2003). In fact, there is a consortium of organisations in Wales which are looking to ensure data science and AI are at the forefront of Welsh research, known as the Wales Data Nation Accelerator (WDNA) (Welsh Government, 2022). Additionally, the Advanced Manufacturing Research Centre Cymru (AMRCC) are looking at ways to

incorporate AI to solve manufacturing challenges for the Aerospace and Food & Drink sectors (Richardson, 2022).

AI Adoption Challenges for Welsh Manufacturing SMEs

Despite the investment and effort put in to elevate AI research in Wales, there is a fundamental challenge in getting the Welsh manufacturing sector to adopt AI. A poll by MADE Cymru revealed 37% of respondents felt that the biggest adoption challenge for digital technologies is cultural/resistance to change, followed by lack of knowledge (MADE Cymru, 2022). An AI workshop was held at Swansea University in 2019 (Swansea University, 2019) which revealed the adoption of AI is hindered by the availability of data, misperception of AI applications and high system costs (Dwivedi *et al.*, 2021). There is little evidence on the challenges Welsh manufacturing SMEs face regarding the adoption of AI, however a cross link can be drawn against European SMEs. The current challenges entail intimidation by large organisations adopting AI, access to data, lack of funding, high costs to implement a customised AI application and lack of digital skills to develop AI applications (Watney and Auer, 2021). A research paper regarding the empirical study into ML for SMEs found that SMEs are less likely to have ML technologies because they struggle to identify use cases for ML applications and have poor data quality (Bauer, van Dinther and Kiefer, 2020).

Introduction into TinyML

AI is a relatively new field for the Welsh manufacturing sector and requires practical intervention to reduce the barrier to entry for Welsh SMEs. The globe has seen significant advancements in miniaturizing ML algorithms and increasing processing speeds in electronic boards (Stewart, 2020). This ultimately paves the way for a new field of study called **Tiny Machine Learning** (TinyML), or also known as Embedded Machine Learning. TinyML focuses on combining machine learning and embedded systems (IoT devices) to provide applications which are executed on small, real-time, low-power and low-cost embedded devices (Janapa Reddi *et al.*, 2022a). An illustration is shown in **Figure 2**.

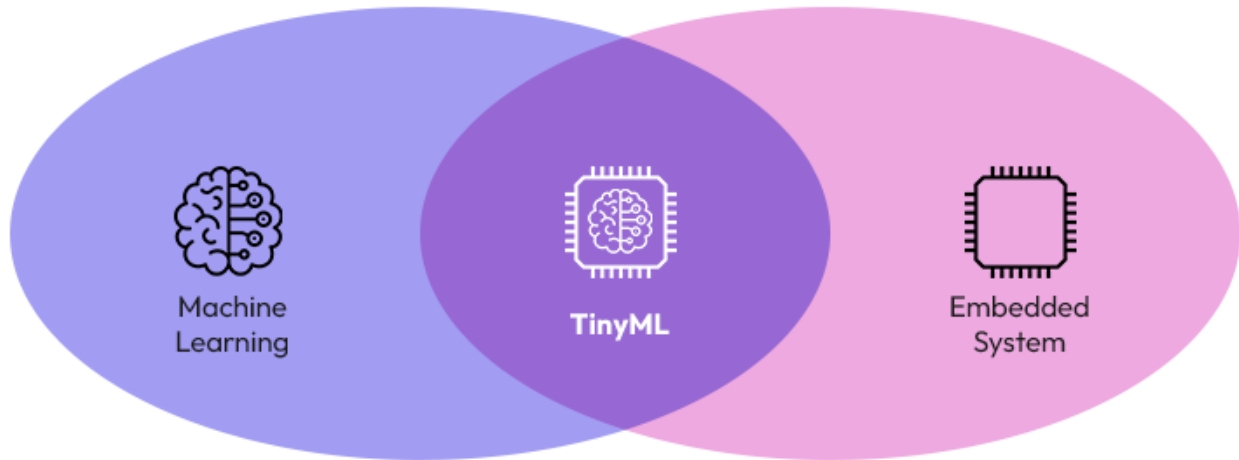


Figure 2: TinyML

TinyML looks to provide AI applications which are typically run in the cloud or edge (high performance computers) at a fraction of the cost, consuming less energy and requiring no internet connection. A typical example of TinyML is speech recognition, to recognise key words which allow users to log actions or people with certain disabilities to switch equipment on or off (Wong *et al.*, 2020).

The importance of research into TinyML cannot be underestimated as it takes advances of existing IoT devices and provides intelligence without needed to purchase additional hardware. According to a forecast by ABI Research, by 2030, around 2.5 billion smart devices will reach the market through TinyML (ABIresearch, 2021). An increasing trend of TinyML worldwide is reflected in **Figure 3** (Google Trends, 2022). Additionally, a Scopus trend shown in **Figure 4** illustrates the increasing publications around TinyML (Scopus, 2022).

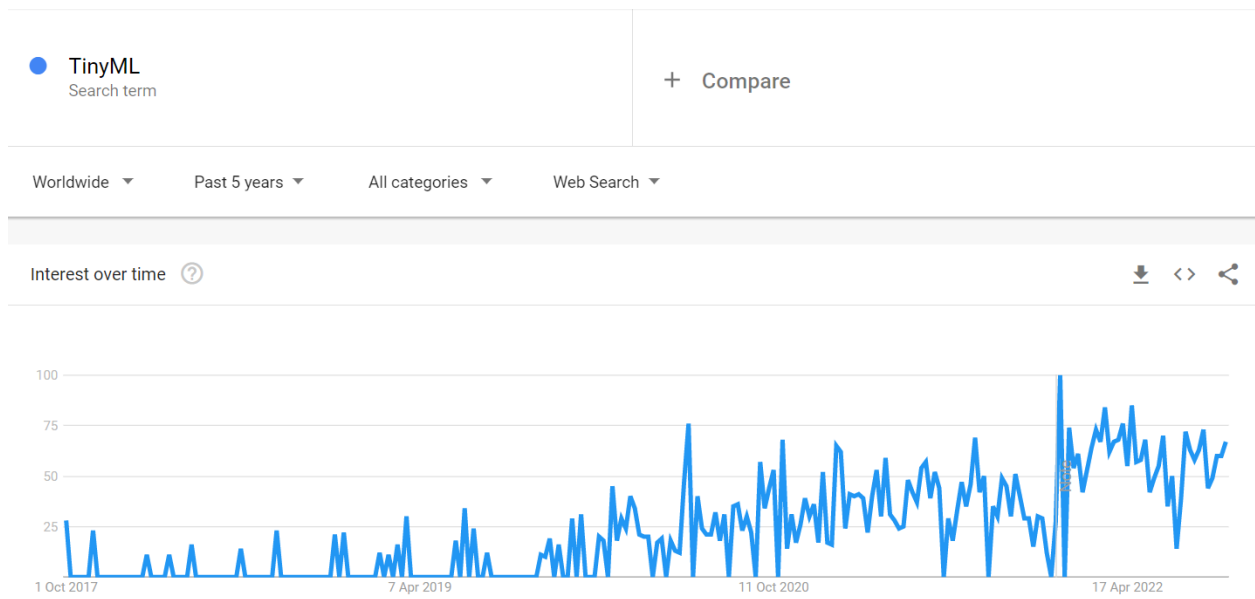


Figure 3: TinyML Interest of Time using Google Trend

Documents by year

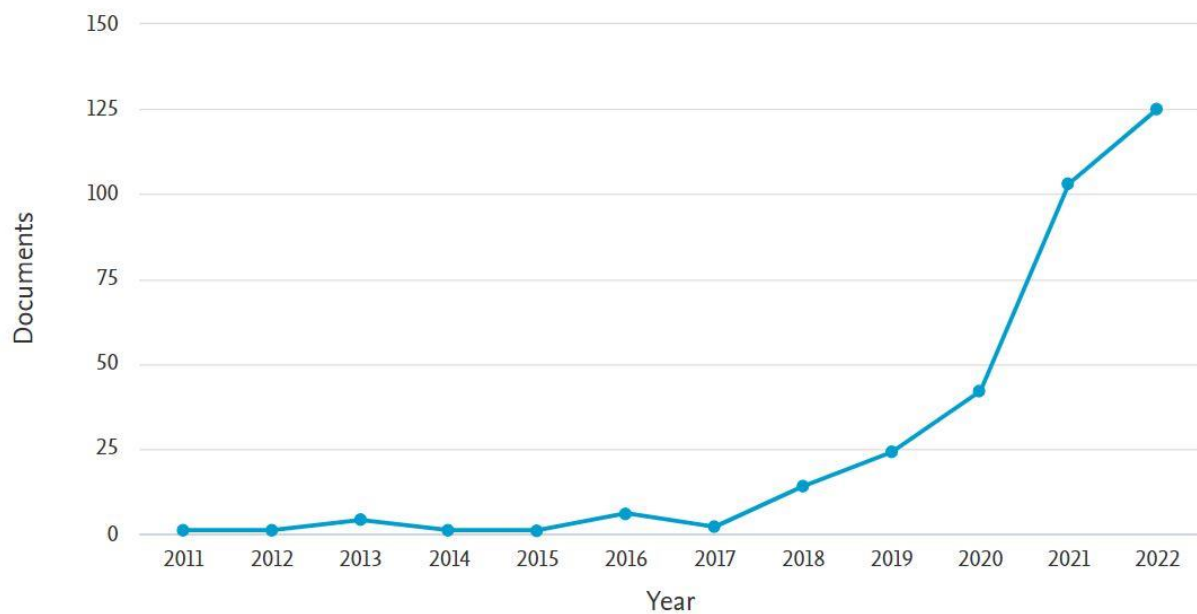


Figure 4: Documents published relating to TinyML, search term of "TinyML" OR "Embedded Machine Learning" OR "tiny ml" OR "Tiny Machine Learning"

Research Problem

There isn't enough concrete evidence online to suggest that Welsh manufacturing SMEs are adopting AI. The challenges stem from high implementation costs, lack of digital skills and lack of use cases. This is where the entire ecosystem involving TinyML may look to reduce the barrier of entry into the adoption of AI and bring value to Welsh manufacturing SMEs through the use of intelligence devices to improve productivity, quality and processes. The hype around TinyML is huge and it is clear that there are plenty of literature available around the use cases involving TinyML i.e. anomaly detection, people counting, speech recognition etc... (Ray, 2022) which are predominately used by non-commercial users such as hobbyists and researchers.

Given the fact TinyML is a new field of study, it is surprising to see the rapid advancement in the solutions available to deploy this technology i.e the hardware, software and methodology.

However, there isn't much evidence available suggesting that TinyML is being used in an industrial sense (Higginbotham, 2022), in particular, the manufacturing sector. More concerningly, there isn't a research paper that explores the applicability of TinyML for the Welsh manufacturing sector in terms of use cases and its impact on businesses. On a broader sense, there is a lot of market hype but a lack of coherent use cases to prove whether TinyML can bring value to manufacturing processes. This leads us to question whether TinyML is feasible for industrial applications.

Therefore, this research proposes an investigation into the applicability of adopting TinyML for Welsh manufacturing SMEs. There are two parts to this research, firstly reviewing where TinyML can be implemented in the manufacturing sector and secondly, conducting a representable manufacturing use case using TinyML through commercial (low-no code) platforms and low-cost technology.

The importance and outcome of this proposed research will dictate whether TinyML is suited to be a viable technology entry for Welsh SMEs. The potential impacts of this research will help researchers and technology providers navigate their focus on TinyML towards the

manufacturing sector, but also enable the company to begin showcasing practical applications of TinyML for the Welsh manufacturing sector, at the company's facility. This will help customers gain an appreciation of the technology and methodology used but also aid customers in generating further ideas on where TinyML can be used for their own firm. This leads the company to potentially generate more business by securing further research projects.

Literature Review

Manufacturing Challenges for SMEs

Given the fact that SMEs are struggling to adopt AI, a step back needs to be taken to understand the core manufacturing issues which SMEs are facing on a day-to-day basis, this piece of literature will help signify where TinyML could intervene to ease the challenges faced by SMEs.

Firstly, preserving quality in production lines is an issue for SMEs in manufacturing (Abel-Koch *et al.*, 2019). This is linked due to human errors on the line (distraction or fatigue) (Torres, Nadeau and Landau, 2021) or machine failures (Jarrett, 2022). This creates unwanted costs for businesses to rework or dispose of defected products.

Labour shortages is a big issue across the UK manufacturing sector, for example, the food & drink industry which originated from the rise of Brexit and the recent Covid-19 pandemic (Grant Thornton UK LLP, 2021). The younger workforce in the UK have grown up with digital technology, hence there is an expectation that SMEs should provide jobs of a digital capacity, but SMEs are unable to fulfil those jobs due to the cost and complexity of implementing digitisation solutions (onrobot, 2022). It is found that 49% of UK manufacturing SMEs believe recruitment is hindering growth of their business (SWMAS, 2021). In Wales, SMEs currently faces a challenge to get access to skills to help run their business. An employer skills survey was conducted in 2019 and it was found that labour intensive, middle skill and high skill roles were hard-to-fill (Department of Education, 2020). A MAKE UK start up report states that the barriers SMEs face is securing and retaining skills and adopting technology to improve competitiveness (Khan, 2021).

Resistance to change is a challenge posed on manufacturing SMEs. A case study conducted by Cardiff University found that a particular SME was resistance to change when it came to the adoption of advanced technology (Walters, 2008). Another case study using

a holistic assessment tool to assess an SME manufacturer's level of digitisation found that the fear of change was the main factor in adopting digital technology (Ottesjö *et al.*, 2020).

Supply chain is another challenge SMEs face. About 96% of UK SME manufacturers are struggling with price changes within the supply chain (SWMAS, 2021). Which is being caused by the availability of raw materials and transportation costs. Due to this fact, it has been difficult for SMEs to grow their business.

Due to the rise of inflation, the energy costs have risen which has limited SMEs ability to scale up. SMEs are constantly coping with price fluctuations and are in need for energy strategies to conserve electricity (Datahoney, 2021).

In terms of digital solutions which SMEs would like to see solve their challenges, a UK manufacturing workshop involving 128 SMEs of all types produced in a paper detailing the solutions areas that SMEs would prioritise first. The results are shown in **Figure 5** (Schönfuß *et al.*, 2021).

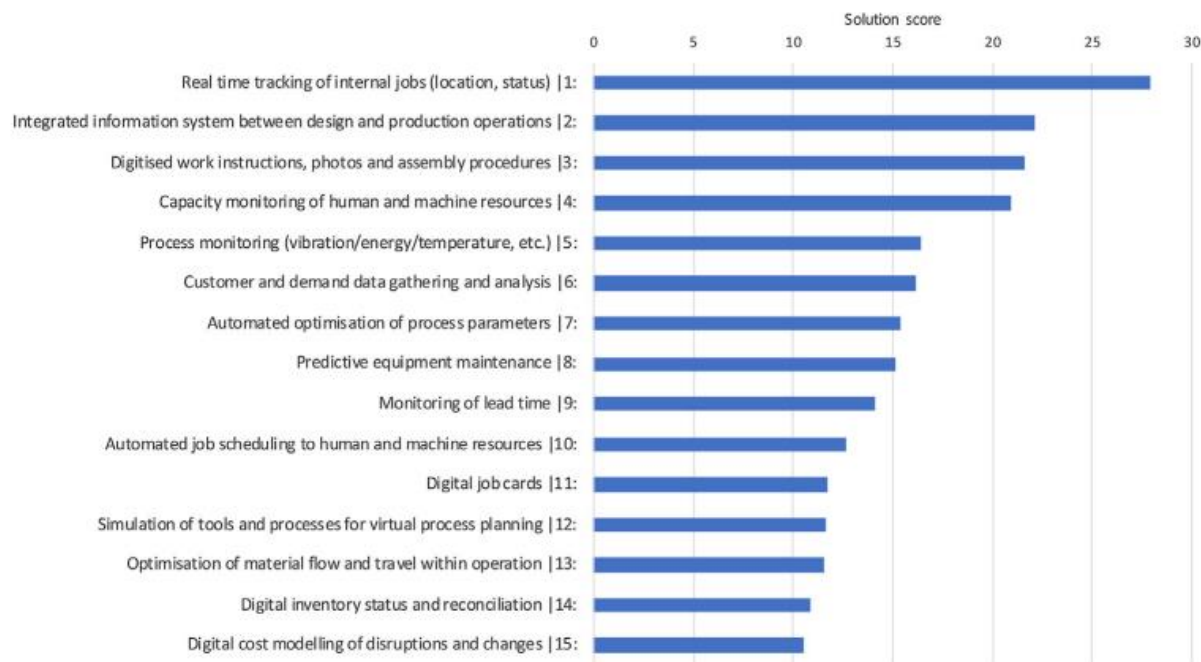


Figure 5: Results of important solution areas viewed by manufacturing SMEs

The ability of having a real time tracking system of internal jobs and digitalised work instructions is considered most important as SMEs typically manufacture in low volumes,

high variety. A lot of the current processes involve humans to keep track of products flowing throughout the factory and manually assembly a product based off memory (Schönfuß *et al.*, 2021). It's interesting to note that the requirement to implement process monitoring and capacity monitoring of human and machine resource is quite high, this is of interest because TinyML can help pose as a solution at a low implementation cost.

It is important to note that a lot of literature around manufacturing challenges for SMEs are addressed at a high-level and do not give real examples of problems. More importantly, it is noticed that there is a lack of literature or evidence detailing which individual manufacturing processes or operations are an issue for Welsh SMEs in manufacturing i.e. exactly what is going wrong in production, quality or inventory? This signifies the importance of this research to identify those challenge areas SMEs face, which will directly help link where TinyML can be applied to solve an industrial problem.

Applications of TinyML

The applications of TinyML are still be uncovered as the technology is new. However, within the research field, researchers have managed to find several applications and the common outcome between all applications shows that TinyML can reduce the workload of the user. The section briefly talks about the different applications in which TinyML is introduced.

Keyword Spotting (KWS) – KWS is a speech recognition technique that is able to identify specific words from a short voice recording. The machine learning model in a TinyML system is able to identify predefined keywords which could be used to trigger a signal to 'wake up' or 'power down' a larger system which does a more complex task (Grau and Centelles, 2021). Ultimately, the keyword acts as a trigger which can in-turn help a connected system perform a larger task (i.e. opening a door, moving a robot or commanding autonomous guide vehicles to go to specific missions). Grau and Centelles developed a TinyML system to recognise the words; vermell, verd and blau (red, green and blue in Catalan) and the system successfully proved its functionality over several tests. (Grau and Centelles, 2021).

Visual Wake Word – This is an extension of KWS, where the TinyML model can spot specific objects or person within an image and creates a trigger response if it sees the object of interest. An example of this application in action can be inspired from the company Bird Buddy. This company designed a smart bird feeder with a built-in camera that can recognise if birds are present at the feeder location, if so, the TinyML system sends a notification to the user (Grau and Centelles, 2021). The caveat to this application is that TinyML systems are not capable in recognising detailed, complex objects (i.e. defects in turbine blades) as it requires significant RAM usage and surpasses the capability of existing low-cost MCU hardware in the market.

Anomaly Detection – TinyML can be used to identify unexpected events. Anomaly detection usually follows an unsupervised learning approach where the machine learning model learns to discover patterns on unlabelled datasets (Grau and Centelles, 2021). There are a lot of industrial systems which already use anomaly detection features, hence making this application a likely candidate for TinyML systems. Examples of anomaly detection can be directly related to manufacturing, where the system is used to detect malfunctions on factory machines or motor bearings (Banbury *et al.*, 2020). This is possible through small sensors (IMU) being attached to the machine which is persistently monitoring vibration, sound or temperature (Grau and Centelles, 2021).

Environmental Sensing – TinyML systems can be used to classify environmental conditions through sounds, vibration and colours. For example, it could be used to identify if people are cutting down trees in an unauthorised location or animals present within an area through sounds. A paper proved this using an Arduino Nano 33 BLE sensor and created a TinyML system which is capable of discriminating several range of sounds, colours and vibrations patterns (Delnevo *et al.*, 2021). Although it is noted that the researcher conducted the experiments in a lab and not outside conditions, it remains unseen whether the TinyML system can replicate the same level of accuracy outdoors. Another researcher managed to create a TinyML system in the form of a wearable device to validate various types of leaves. The result of this research led to a developed TinyML system which could classify diseased leaves with 90% accuracy (Silva *et al.*, 2021). Another example is a researcher using the Arduino Nano 33 BLE Sense to perform time-series forecast of greenhouse temperatures

(Alati *et al.*, 2022). This example looked at real imported datasets but has not tested the system with real-time data, thus lacks evidence that it will perform accurately when the device is exposed to outdoor conditions.

Transport – Given the rise of embedded algorithms, research is being conducted to see if TinyML systems can benefit vehicles to make them more ‘smart’. There could be possibility that it could be used to understand the amount of air pollution a vehicle provides. Andrade et al conducts an experiment to provide a solution to compute CO2 emissions from vehicles by directly reading their OBD-11 interface. The results were promising in terms of accuracy but the challenge remains in trying to mature the solution, which requires multiple vehicles to be tested on (Andrade *et al.*, 2022).

Health – TinyML has already been proven to work in industry through Fitbit or smart wearable devices which measure vitals such as heart rate, blood oxygen concentration, activity etc.. (Shafique *et al.*, 2021). This can provide the user an update on their health which allows them to take insightful decisions on how to improve their health. It is possible to use TinyML systems to visually monitor patients’ actions or identify the type of medicine is required. This allows cake-takers or doctors to take effective measures in ensuring the patient is treated well.

Security and Surveillance – TinyML can be used for security purposes to detect if a person is present in an unauthorised location (Shafique *et al.*, 2021). Additionally, it could be used to identify objects which could potentially cause harm to a scene i.e. knives, guns, sharp objects etc.

Industry – TinyML, although not proven in industry as of yet, it could be used for quality assurance (Shafique *et al.*, 2021), to determine whether a product is good or bad i.e. if a Cherrybake well is cracked or not.

Supply Chains – Quite recently in the Edge Impulse Imagine Conference (Edge Impulse, 2022b), a company is looking to use TinyML to count the pallets being shipped off to customers and being returned by the same customer. The system would then report back the missing number of pallets to help the company identify the charges needed to be imposed on the customer.

An overview which maps the input data type with TinyML use cases is shown in **Figure 6** (Banbury *et al.*, 2020). It also illustrates the type of neural network is needed to perform the use case and may be a good baseline to help conduct the experiment in this proposed research.

INPUT TYPE	USE CASES	MODEL TYPES	DATASETS
AUDIO	AUDIO WAKE WORDS CONTEXT RECOGNITION CONTROL WORDS KEYWORD DETECTION	DNN CNN RNN LSTM	SPEECH COMMANDS (WARDEN, 2018A) AUDIOSET (GEMMEKE ET AL., 2017) EXTRASENSORY (VAIZMAN ET AL., 2017)
IMAGE	VISUAL WAKE WORDS OBJECT DETECTION IMAGE CLASSIFICATION GESTURE RECOGNITION OBJECT COUNTING TEXT RECOGNITION	DNN CNN SVM DECISION TREES KNN LINEAR	VISUAL WAKE WORDS (CHOWDHERY ET AL., 2019) CIFAR10 (KRIZHEVSKY ET AL., 2009B) MNIST (LECUN & CORTES, 2010) IMAGENET (DENG ET AL., 2009) DVS128 GESTURE (AMIR ET AL., 2017)
PHYSIOLOGICAL / BEHAVIORAL METRICS	SEGMENTATION FORECASTING ACTIVITY DETECTION	DNN DECISION TREE SVM LINEAR	PHYSIONET (GOLDBERGER ET AL., 2000) HAR (CRAMARIUC, 2019) DSA (ALTUN ET AL., 2010) OPPORTUNITY (ROGGEN ET AL., 2010) UCI EMG (LOBOV ET AL., 2018)
INDUSTRY TELEMETRY	SENSING (LIGHT, TEMP, ETC) ANOMALY DETECTION MOTOR CONTROL PREDICTIVE MAINTENANCE	DNN DECISION TREE SVM LINEAR NAIVE BAYES	UCI AIR QUALITY (DE VITO ET AL., 2008) UCI GAS (VERGARA ET AL., 2012) NASA'S PCOE (SAXENA & GOEBEL, 2008)

Figure 6: Mapping of Input data type to TinyML use cases to NN model Types

Benefits and Challenges with TinyML

TinyML is a new research field, thus despite the hype, researchers have found several challenges associated with this area of research. This section will cover the current benefits and challenges associated with TinyML.

The benefits associated with TinyML systems is shown in **Table 1**.

Benefits	Justification
Privacy	Data can be processed at the edge and is not exposed to the wider internet. This makes it less vulnerable for hackers to intercept to steal the data (Ren, Anicic and Runkler, 2021).
Latency	Since the inference takes place at the edge, decisions are made by the system in real-time, which makes it ideal for safety critical applications (Ren, Anicic and Runkler, 2021).
Responsiveness	Since the inference of the machine learning model is offline, there isn't a need to input extra energy to send the data to the cloud and then send the results back to the device (Sudharsan et al., 2021). This creates a delay in terms of inference response time. With the inferencing taking place at the edge, this reduces the response time to provide the result.
Cost	The most obvious benefit of implementing TinyML systems is cost, due to the advancement in electronic chips, embedded boards are now becoming

	easier to buy. Embedded boards which are capable of TinyML are generally cheaper when used over time compared to a cloud-based solution or data centre solution.
--	--

Table 1: Benefits of TinyML systems

The challenges associated with TinyML systems is described in **Table 2**.

Challenges	Justification
Energy Efficiency	Using a neural network requires power and to retrain a model it requires the model to be sent to a larger computing hardware source, thus this process still requires more power than necessary (Ren, Anicic and Runkler, 2021).
Deployment	Sending the data to a cloud service to retrain the model and then redeploy the model back into the hardware is a challenging task due to the isolation the system is in during industry (Ren, Anicic and Runkler, 2021). The methodology of deployment is a key challenge to understand the architectural considerations and metrics to assess such systems (Shafique et al., 2021).
Scale	Updating the model for one device is easy but doing it for large number of devices is an overwhelming task especially if the devices are in remotes locations in industry (Ren, Anicic and Runkler, 2021).

Data Storage	Since TinyML is implemented with MCU embedded devices, these devices have limited possibility to store field data. Thus, large data can only be stored in a data centre which is expensive (Ren, Anicic and Runkler, 2021).
Data Drift	Environments in which devices are operating in are constantly changing, thus it is a challenge to ensure the ML model performance stays up to date to suit the changing environment (Ren, Anicic and Runkler, 2021).
Chipset security	Embedded devices are typically development boards meaning it is quite easy to trick the device and download malware from an unknown server (Sudharsan et al., 2022).
Network connection	During over-the-air updates to embedded devices, there is a risk that an uninterrupted network connection can happen which may corrupt the files to be installed onto the hardware (Sudharsan et al., 2022).
Maturity of Software and datasets	A research paper states that finding software products that belong to embedded devices was a challenging task due to the workflow or concept of TinyML not being properly defined (Sipola et al., 2022). However, one can argue that this is no longer the case thanks to vendors such as Edge Impulse and Neuton.AI who are

	looking to provide software to enable users to build TinyML systems. The lack of datasets also makes it difficult for researchers to evaluate this technology and thus there is a requirement for more open-source datasets to promote benchmarking of TinyML systems (Banbury <i>et al.</i> , 2020).
Robustness and Reliability	Shafique <i>et al</i> highlights that for tiny machine learning at the edge, the maturity has not been achieved and requires further datasets and stimulation through more use cases to enable society to accept the technology (Shafique <i>et al.</i> , 2021).
Memory	There is a limitation in the memory embedded devices can have to process AI algorithms and still requires further research to understand the most efficient way to minimise device memory usage whilst processing the data onboard the device (Sudharsan <i>et al.</i> , 2021). A research paper has attempted to solve this issue through a differentiable NAS (neural architecture search) search which automates the design of neural networks to fit around the application (C. Banbury <i>et al.</i> , 2020).

Table 2: Challenges of TinyML systems

This research is looking to tackle two challenges, the maturity in software and datasets and robustness and reliability. By conducting an experiment, this research will contribute to providing open-source datasets for researchers to experiment with and provide an evaluation as to how robust and reliable off-the-shelf TinyML technologies are.

Hardware, Software and Models for TinyML Applications

TinyML is the intersection between machine learning and microcontrollers, it is the miniaturisation of neural networks on MicroController Units (MCUs) to be able to detect, classify, recognise and predict on ultra-low power boards (Sudharsan et al., 2021), similarly to regular neural networks on larger computing systems (Su et al., 2022). Note for clarification, microcontrollers, embedded boards, embedded devices are loosely termed together which has similar meaning.

TinyML systems comprises of the hardware (boards), software, sensors and user interface and can be in the form of a wearable device, pocket device or small bespoke kits.

To build TinyML applications, there are two ways in which a developer can approach this, via frameworks to write code to deploy onto microcontrollers or software which is able to automatically carry out the end-to-end methodology of training and deploying the model onto the microcontrollers. The commonly used frameworks which are used to build TinyML applications are described below (Ren, Anicic and Runkler, 2021), (de Prado et al., 2019).

- CMSIS-NN from Arm
- TensorFlow Lite for Microcontrollers from Google – Examples are shown by Grau and Centelles which use an Arduino Nano 33 BLE Sense and an Espressif ESP32 board to create basic TinyML applications (Grau and Centelles, 2021).
- Caffe2 - a lightweight deep learning framework focusing the deployment of AI on mobile devices. Caffe2 can be accessed and utilised in Python.
- Android Neural Networks API (NNAPI) - A C API designed for deployment on Android devices. This API cannot be used alone to create TinyML applications, it only optimises models from Tensorflow or Caffe2 which can be deployed onto Android devices.
- Arm Compute Library (CL) - A framework that collects low-level functions optimized for Arm CPUs and GPUs. This library is built help with image processing, computer vision, and machine learning tasks. This framework is useful for boards with have Cortex Arm chips.

- Nvidia TensorRT - Built on CUDA and it provides capabilities to optimize a trained model for embedded devices. TensorRT is widely compatible with neural network models trained in major frameworks such as TensorFlow, Pytorch, Keras etc.

These frameworks ultimately optimise neural networks in such a way to work on microcontroller chips without overloading the boards.

Next the software which automates end-to-end TinyML applications are shown in **Table 3** (Sawyer, Croome and Jongboom, 2022). This will be a key for this research proposal as SMEs are looking for an easy way of adopting technology, thus by using available software, the businesses can avoid coding and instead focus on the application they wish to build.

Software	Description
Edge Impulse	A development (low-no code) platform for embedded machine learning which enables the user to connect to a device to capture and label data, perform feature engineering, train and optimise the model, deploy the model – all within a few clicks.
Neuton.AI	A no-code platform that enables you to build TinyML applications on 8,16,32 bit MCUs. It offers similar capabilities to Edge Impulse but more tailored for the cheaper end of the microcontroller market.
SensiML	Another platform like Edge Impulse that provides a complete workflow for building TinyML applications.
OmniML	OmniML focuses on accelerating models into production, but operates very similarly to SensiML and Edge Impulse.
Imagemob AI	Similar capabilities to Edge Impulse.

Table 3: TinyML Software to help build applications (low-no code platforms)

For code development, Python and C++ is the typical language used given its compatibility with MCU boards. Pycharm and Jupyter Notebooks are coding environments which are used to help users develop applications (Grau and Centelles, 2021).

In terms of TinyML models and overcoming the limitation of boards storage capacity, a researcher proposed a novel system called TinyOL (TinyML with online-learning) which learns from datasets incrementally and does not require the need to store historical data (Ren, Anicic and Runkler, 2021). From this, the model is always up to date and does not require additional work to account for concept drift. The researcher created two use cases to show this new model but requires further work in maturing this model for wider adoption. There are other models which have been built from research and tailored to suit different TinyML applications. These models are shown in **Figure 7**. It is possible to use these models in conjunction with a framework to quickly build a TinyML application for this research project but requires extensive testing to understand the robustness of these models.

Task: Model Name	Score	.tflite (KB)	.h (KB)
Recognize Gestures: MagicWand	0.67 (Acc)	19	118
Visual Wake Words: MicroNet S-L	0.76-0.82 (Acc)	273-529	1689-3267
Speech Recognition: Wav2letter	0.0783 (LER)	22600	143421
Keyword Spotting: DNN S-L	0.82-0.86 (Acc)	82-491	508-3029
Keyword Spotting: CNN S-L	0.91-0.92 (Acc)	75-492	436-3029
Keyword Spotting: MicroSpeech	0.62 (Acc)	18	112
Image Classification: MobileNet v2	0.69 (Acc)	3927	24215
Anomaly Detection: MicroNet S-L	0.95-0.96 (AUC)	246-452	1523-2794

Figure 7: Examples of TinyML Models which are dedicated for various applications, the accuracy of the model's performance is also shown

An example of the hardware which are suitable for TinyML applications is shown in **Figure 8** (de Prado *et al.*, 2019). Note that the reason these boards work because there are frameworks (as described from above) which considers the optimization of neural networks for Cortex ARM chips.

Name	Processor	Flash (MB)	SRAM	Clock (MHz)
B1: Teensy 4.0	Cortex-M7	2	1 MB	600
B2: STM32 Nucleo H7	Cortex-M7	2	1 MB	480
B3: Arduino Portenta	Cortex-M7+M4	16	8 MB	480
B4: Feather M4 Express	Cortex-M4	2	192 KB	120
B5: Generic ESP32	Xtensa LX6	4	520 KB	240
B6: Arduino Nano 33	Cortex-M4	1	256 KB	64
B7: Raspberry Pi Pico	Cortex-M0+	16	264 KB	133

Figure 8: Examples of typical TinyML hardware found from research

TinyML Deployment Methodologies

The method to deploy TinyML systems is not coherent according to research. It generally follows the same methodology of; collect data, pre-process data, label data, train model, test model and deploy model. Deploying TinyML systems require some level of expertise in both software and hardware in order to be successful (de Prado et al., 2019)

The workflow for running AI models on microcontroller hardware has not yet found a best practice. A researcher explains that to get the model onto the hardware there are three possible ways (Sipola et al., 2022).

- 1) Use a runtime that loads the model data.
- 2) Using a transcompiler to compile model to C or C++ code to be implemented onto the hardware
- 3) Using a compiler that compiles the model to a library and is linked to the hardware

More recently, it is now commonly seen that the majority of people who are involved in creating TinyML applications are using Edge Impulse due to the ease of use and the fact that it does not require deep knowledge of machine learning or embedded systems to work. A lot of the methodology within that platform is automated (Grau and Centelles, 2021).

TinyML Benchmarking

There have been a few research papers looking at benchmarking TinyML hardware against each other to understand which board provides a good cost-performance ratio. One example is a piece of research looking at benchmarking 7 different boards with 5 different datasets (Sudharsan et al., 2021). The benchmark test is shown in **Figure 9**.

MCUs, DATASETS, NNS CHOSEN FOR TINYML BENCHMARK.		
	Board Name	Processor, Flash (MB), SRAM, Clock (MHz)
MCUs	B1: Teensy 4.0	Cortex-M7, 2, 1MB, 600
	B2: STM32 Nucleo H7	Cortex-M7, 2, 1 MB, 480
	B3: Arduino Portenta	Cortex-M7+M4, 2, 1MB, 480
	B4: Feather M4 Express	Cortex-M4, 2, 192KB, 120
	B5: Generic ESP32	Xtensa LX6, 4, 520KB, 240
	B6: Arduino Nano 33	Cortex-M4, 1, 256KB, 64
	B7: Raspberry Pi Pico	Cortex-M0+, 16, 264KB, 133
Name: Feature dimension, Class counts		
Datasets	D1: Iris Flowers: 4, 3	D6: Breast Cancer: 30, 2
	D2: Wine: 13, 3	D7: Texture: 40, 11
	D3: Vowel: 13, 11	D8: Drive Diagnosis, 48, 11
	D4: Silhouettes: 18, 4	D9: MNIST Digits: 64, 10
	D5: Anuran Calls: 22, 8	D10: Human Activity: 74, 6
Name: Topology		
Fully	FC 1x10: 1 layer with 10 neurons	
Connected	FC 10+50: 1 st layer with 10 neurons, 2 nd with 50	
NNS	FC 10x10: 10 layers, with 10 neurons in each layer	

Figure 9: Benchmarking Test

The results of this test showed that the Teensy 4.0 board is the fastest at providing inference results for the price stated. However, across all the boards, only 0.4-1.6% variation was shown in terms of accuracy for inferencing. The problem with this paper is that it does not tackle onboard functionality to see how well it copes in an industrial environment and with real data, so it does really tell us anything about whether the boards are simply capable of being used in industry. Another short paper looked at highlighting the need of benchmarking TinyML systems but with open-source datasets to understand the limitations of the boards (Banbury *et al.*, 2020). Both authors only focus on the limitation of the hardware in terms of

neural network performance but not the whole ecosystem that surrounds it, which is more important to understand from a business perspective on whether the TinyML system is capable of performing well in industry and provide useful insights to benefit the workforce.

Research Aims and Objectives

The aim of this research is to investigate the applicability of adopting TinyML for Welsh manufacturing SMEs. It will ultimately look to prove whether SMEs should be seriously considering the adoption of TinyML systems to improve their manufacturing processes.

Based on the initial literature review, there it is clear that there are gaps in research regarding the identification of manufacturing challenges for Welsh SMEs as well as the viability of integrating TinyML systems in an industrial application and understanding its reliability and accuracy in terms of functionality. Additionally, it is not clear where else TinyML can benefit manufacturing processes apart from predictive maintenance and anomaly detection. Thus, this research will also need to explore other possibilities on where TinyML can be introduced to improve manufacturing operations. The objectives of this research are as follows:

- 1) Identify the key manufacturing challenges for Welsh SMEs.
- 2) Conduct an expanded literature review to explore other areas where TinyML systems may help manufacturing operations.
- 3) Map TinyML technologies and solutions to the identified Welsh SME manufacturing challenges.
- 4) Conduct an experiment for an anomaly detection of an industrial conveyor in an industrial environment using existing solutions and assess the reliability, functionality and robustness of the chosen solution.

Research Scope

- 1) Only one TinyML solution shall be explored and implemented in the experiment, for the interest of time.
- 2) The research around finding key manufacturing challenges for SMEs is only limited to Wales.
- 3) This research is not looking to build products or develop new machine learning algorithms, it shall be limited to using off-the-shelf technologies and software.

Proposal

Methods

For this research, a mixed method approach of a survey and experiment will be conducted. The survey aim is to collect information on the key SME manufacturing challenges in Wales, whilst the experiment aim is to test the ease of setup and reliability of TinyML systems in an industrial setting.

Interviews and focus groups are not appropriate for the collection of key SME manufacturing challenges due to the interest of time as the idea is to collect information from as many businesses as possible to provide a strong evidence base. Ethnography is also not pursued in this research as every manufacturing business is unique in terms of challenges, hence observing one company may have different results compared to another due to the differences in manufacturing process and products. Thus, trying to obtain the key SME manufacturing challenges would be biased towards a single company.

Survey

The survey will be conducted through an online based approach through SurveyMonkey (SurveyMonkey, 2022) thanks to its ease of use. In terms of survey reach, the AMRC has a contact list/network of Welsh businesses which they have interacted in the past. Furthermore the AMRC hosts cluster events where Welsh business attend. Thus, the survey can be distributed through those contacts to collect a good set of results. The aim is to achieve at least 30 responses.

There will be 5 questions asked to obtain knowledge around the key SME manufacturing challenges. Note that the idea of the survey is not to introduce TinyML to SMEs as it is most likely the businesses won't have sufficient understanding to answer questions about it accurately. The 5 questions the survey will ask is:

- 1) What manufacturing sub-sector do you belong to?
- 2) How many employees do you have in your business?
- 3) What are the main challenges for your business (tick all that apply and provide a sentence on what your challenge is about)
 - a. Raw Ingredient Procurement
 - b. Raw Material Storage
 - c. Production Scheduling
 - d. Safety
 - e. Product Production
 - f. Product Packaging and Palletizing
 - g. Quality Control
 - h. Finished good warehousing
 - i. Customer Logistics
 - j. Customer Ordering
- 4) Have you engaged with Industry 4.0 technologies before? I.e. IoT, AI, AR/VR, smart robotics
- 5) What is the main barrier (if any) stopping you from adopting industry 4.0 technologies?

For ease of mind, the survey will not ask for emails or names, to ensure confidentiality. Question 1 and 2 will help to understand the type of business the respondents run and will hopefully identify any cross sub-sector link in terms of manufacturing challenges. Question 3 will enable the respondents to identify the type of challenge they are facing but will go a step further in providing an opportunity for the respondent to tell us specifically what the challenge is about. This piece of information will enable the research to identify exactly where TinyML may reduce the challenge. Question 4 will seek to understand whether Welsh SMEs have encountered with industry 4.0 technologies, this will aid the research in evaluating the level of awareness SMEs have. This would then help researcher's pitch TinyML at the right level to possibly avoid overwhelming the technology when it is showcased to SMEs.

Experiment

For the anomaly detection experiment, it will be conducted at the AMRC Cymru shopfloor as it contains representable industrial machinery and environment. The application will be about anomaly detection for industrial conveyors, as conveyors are a common item seen across all manufacturing businesses. An image of the conveyor at the AMRC is shown in **Figure 10**. The conveyor consists of; a belt, motor unit, idlers, control unit and gearbox.



Figure 10: AMRC Cymru Food & Beverage Industrial Conveyor

In industry, if conveyors fail or do not operate correctly, there is a possibility that the production line lacks in yield or products are damaged. Taking inspiration from literature by Zimroz and Krol, this experiment will explore the following failure modes (Zimroz and Krol, 2009):

- Belt Slippage
- Seized or misaligned idlers
- Abrasion

For belt slippage, the conveyor belt can be loosened or tightened to replicate a failure. For misaligned idlers, the best way to replicate this failure is to take one or two idlers out of the conveyor to see if the system can recognise the failure, or the idlers will be put to one side of the conveyor to replicate misaligned idlers. To replicate the failure of abrasion, a small 3D printed object will be placed on the underside of the conveyor, such that the belt can ‘rub’ on the object.

To implement a TinyML system to recognise these failures, the deployment methodology shown by Janapa Reddi et al will be used, **Figure 11** illustrates the process (Janapa Reddi *et al.*, 2022b).



Figure 11: AMRC Cymru Food & Beverage Industrial Conveyor

Edge Impulse will be chosen as the software to help implement the TinyML system (Edge Impulse, 2022a). This platform is chosen as it facilitates end-to-end machine learning development. Additionally, since manufacturing SMEs are unlikely to know how to develop code, the case for using Edge Impulse will showcase the difficulty for a user to start creating machine learning applications. According to **Figure 11**, Edge Impulse will carry out all of the steps required to build the application. The pre-processing blocks required for this application data is DSP (Digital signal processing) to ensure that any unwanted frequencies are filtered out. A multi neural network classifier is required for this application to identify when the conveyor is; normal, under abrasion, idler problem and belt slippage.

To identify the failures, three sensors will be used: acceleration, temperature and noise. The sensors will be placed on top of the conveyor motor in a non-invasive manner, near the motor bearings. This placement of sensors is chosen at this location as every failure originates from the load of the motor. Thus, capturing the motor vibration, temperature and noise will enable us to identify when the conveyor is subject to failure. If one sensor is used for classification, then there may be a possibility that the accuracy of the classifications is reduced.

To capture the data, Arduino Portenta H7 with the Nicla Sense ME will be used. The Nicla Sense ME is used due to the industrial sensors embedded onto the electronic board i.e. acceleration (BHI260AP) and temperature (BME688), whilst the Arduino Portenta is used for the onboard microphone. Using these boards in unison is important as the Nicla Sense ME will be capturing and processing sensory data and the Portenta H7 (due to Cortex M7 and M4 cores) will run the inferencing to determine the condition of the conveyor. Arduino claims that Portenta family is intended for industrial applications but has never been proven in reality, hence making this selection a good case to understand whether these embedded boards are up to the task. Additionally, Edge Impulse has official support for the Portenta which allows data collection to be relatively easy through the software.

To collect the data, the Portenta H7 will be connected to a laptop which will run Edge Impulse to collect and label the datasets. This data is held on Edge Impulse's cloud. For this research, sensitive data is not an issue as the point of the research is to share the application across multiple SMEs and help them gain an understanding on TinyML systems. Additionally, an MySQL server (database) will be set up locally on a computer to capture data after the machine learning algorithm is trained, this will be possible through the wireless connectivity from the Portenta H7. The reason for this addition is to train the machine learning model if data drift occurs i.e. when the conditions of the conveyor naturally drifts due to the conveyor being placed in a different environment on the shopfloor.

The workflow to conduct the experiment for the failure modes is illustrated in **Figure 12**.

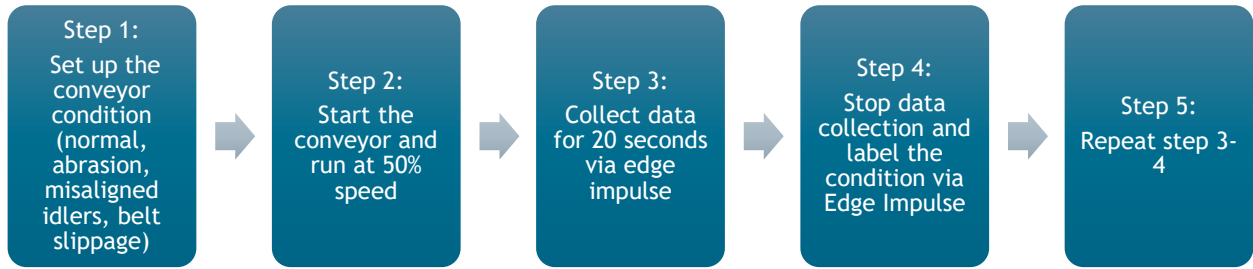


Figure 12: Experiment workflow to capture necessary data

To train machine learning algorithm, the model needs to be trained using a ‘training dataset’ and tested using a ‘testing dataset’. The typically split is 80-20% training to testing. The initial number of datasets to collect per condition will be 100 samples (100x20seconds). In total, that is 400 samples. Thus, 320 samples will be used for training and 80 samples will be used for testing the machine learning model.

The deployment of the model onto the Portenta H7 will be conducted via Edge Impulse. Edge Impulse is able to do this due to its C++ compiler which minimises the neural network to fit within embedded boards. Once this occurs, it will be possible to test the board in isolation.

The methodology of testing the board in isolation for robustness and reliability is to create further experiments and collect the inferencing results. To collect the inferencing results from the Portenta board, a Fast API web framework will be made, which uses REST API connection to send the inferencing results onto the local laptop remotely. The web framework will facilitate the experiment by allowing the user to select the experiment to be conducted to keep track of what tests were made. This requires creation and modification of code on the Portenta device to enable the results to be shown on the web server and a program to run on a local laptop to run the web server. The type of experiments to be conducted is illustrated in **Figure 13**.

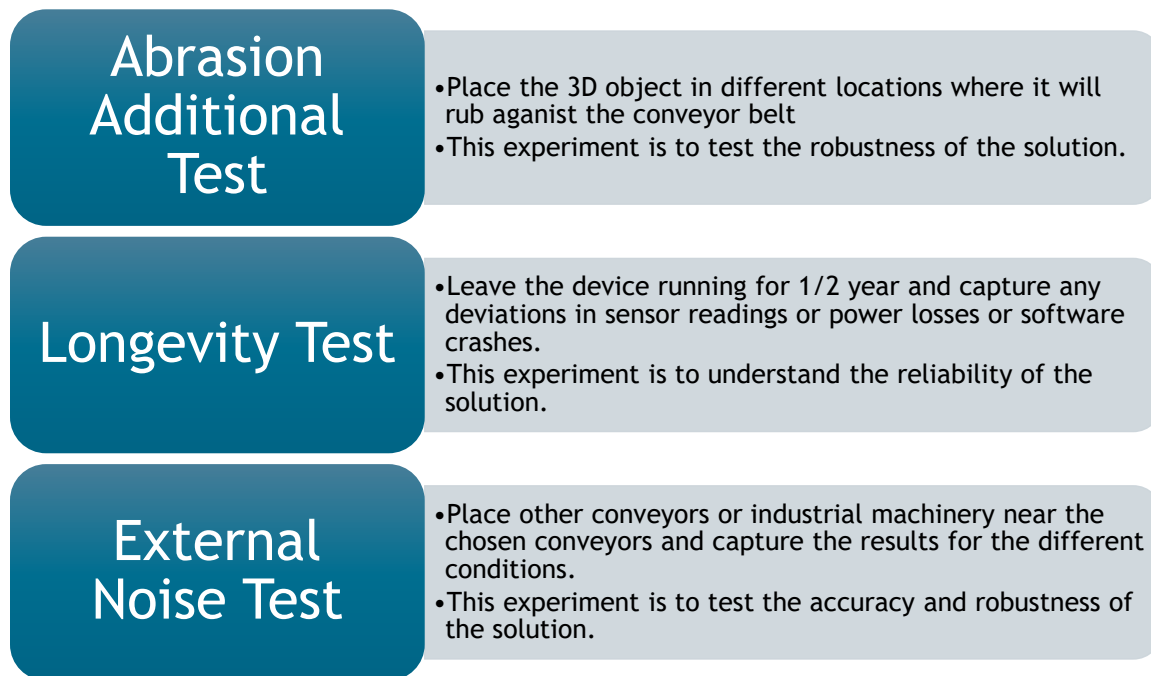


Figure 13: Experiments to test the reliability and accuracy of the TinyML system

Results

The results of the survey will be of a qualitative format. The results will be collected by SurveyMonkey and exported into a Microsoft Excel spreadsheet, where the individual data will be aggregated and shown in a graphical format for reporting.

However, the key is for the results of the survey to be mapped against applications on where TinyML can be introduced (based on extended literature review). This mapping exercise will take place on Lucidchart, which is an online diagramming tool. A mind map can be created which shows the challenges Welsh SMEs face, then stemming from it would be the TinyML application proposal which could reduce that particular challenge. This diagram can be shared in the report for SMEs to view and understand the key technologies, software and implementation considerations to help them get started on adopting TinyML.

For the experiments, as mentioned in the methods section, 3 sensor data is collected from the industrial conveyor: acceleration (vibration), sound and temperature. Upon training the

machine learning model, before it is deployed to the Portenta device, the accuracy of the machine learning model must be at least 85%. If there is a possibility that the accuracy of the machine learning model is not high enough, additional data samples will be collected to further train the model.

In terms of inferencing results (once the machine learning model is implemented onto the Portenta device) Edge Impulse uploads the program onto the device which outputs inferencing metrics (data packets) entailing; classification label, confidence level, time per inference and RAM (memory) usage. The program will also send a data packet (0 or 1) to the mySQL database to identify whether the device is still running in the shopfloor.

To assess the solution's reliability and robustness under each experiment in **Figure 13**. The inferencing metrics will be assessed against several KPIs to indicate to the audience how 'good' the solution really is. **Table 4** shows the KPIs the solution will be measured against. The KPI set is based on existing machine learning solutions implemented in industry.

The report will detail the results from the survey, experiments and the step-by-step methodology used to implement TinyML system on an industrial application. The report will give the SMEs sufficient understanding on what TinyML is capable of and the steps they need to carry out to replicate similar applications.

KPI	Target
Accuracy	>85%
Confidence Level	75-100%
RAM Usage	<95%
Device Measurement Interruptions (due to power loss, crashes or bugs)	0 (1-2 interruptions is allowed as tolerance)

Table 4: KPI to assess the solution in an industrial environment under reliability and robustness tests

Timeline

The estimated timeline for this research is shown in **Table 5** below. The total research project duration is 8 months.

	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8
Survey								
Extended Literature Review								
Survey Creation								
Survey Distribution								
Survey Result Accumulation								
Mapping of TinyML applications against manufacturing challenges from survey								
Experiment								
Procurement of Hardware								
Experiment Set-up								
Data Capture and Labelling for multiple conditions								
Model Training and Optimisation								
Model Deployment (including setup of web server and InfluxDB database)								
Robustness and Reliability Experiments								
Analysis of Experiments								
Dissertation Writing								

Table 5: Research Timeline

Costs

The cost of the research project is shown in Table 6.

Resource Required	Cost (inc VAT)	Qty	Total Cost
SurveyMonkey Standard Monthly Plan	£99/month	7 months	£106
Portenta H7	£100	1	£100
Nicla Sense ME	£70	1	£70
3D Printed ABS Object (for abrasion tests)	£50	2	£100
MySQL Server	Free – Provided by University of Sheffield	1	Free
Laptop	Free – Provided by University of Sheffield	1	Free
Edge Impulse Free Edition	Free (AMRC Cymru can provide a license if needed)	1	Free
Industrial Conveyor	Free – Provided by AMRC	1	Free
Lucidchart Individual Premium License	£85 (annual)	1	£85
Additional Electronics/Equipment	£50	1	£50
Total Project Cost			£511

Table 6: Research Project Costs Excluding Engineer's Time.

Licenses are required for SurveyMonkey and Lucidchart to be able to use the full functionality of the platforms. The hardware can be purchased from RS Components, as it is the trusted UK supplier of electronic components. A budget for additional electronics/equipment is allocated in case of any adaptations needed to the hardware to capture data from the industrial conveyor or any sensor brackets that are needed to be created. The funding source is yet to be confirmed.

Proposal Impact

This research proposal is complementing research in the field of TinyML rather than proposing something novel in order to help build maturity for this technology in the manufacturing sector. The impact of this research proposal will contribute in showcasing practical, industrial use cases for TinyML rather than from a ‘hobbyists’ perspective to show that this technology can be the perfect breakthrough for Welsh SMEs to adopt AI at first instance without being overwhelmed. The report to be generated at the end of the research will help manufacturing businesses understand the steps and technologies required to adopt TinyML and the areas in which it can be applied in a manufacturing context.

In terms of conflicting literature, there isn’t a research paper which directly covers the type of experiment which is being conducted in this research. However this research will lightly touch upon other literature which looked at identifying low-cost machine learning solutions for manufacturing SMEs (Kaiser et al., 2021), but rather go deeper on a technological level. In this research, the exploration of SME manufacturing challenges will not go in-depth (company by company case) compared to a research conducted by Cardiff University, which looked at the impact of advanced manufacturing technologies on SMEs (Walters, 2008). This paper delved into creating a tool to assist SMEs to implement advanced manufacturing technologies. The common outcome from their research in terms of SME challenges was resistance to change (cultural). The outcome of this research proposal should look to contribute to the Cardiff University paper by suggesting low-cost TinyML systems which SMEs can adopt and a practical step-by-step walkthrough (basing on the experiment) to show SMEs how they can adopt TinyML at an affordable, low-cost approach. This will enable SMEs to trial and observe the solutions at their facility rather than commit to heavy initial investment in which they are likely to show resistance to change.

References

- Abel-Koch, D. J. *et al.* (2019) 'Going Digital- The Challenges Facing European SMEs'. Available at: https://www.british-business-bank.co.uk/wp-content/uploads/2019/11/going-digital-the-challenges-facing-european-smes-european-sme-survey-2019_2.pdf.
- ABlresearch (2021) *TinyML Device Shipments to Grow to 2.5 Billion in 2030, Up From 15 Million in 2020*. Available at: <https://www.abiresearch.com/press/tinyml-device-shipments-grow-25-billion-2030-15-million-2020/>.
- Alati, M. F. *et al.* (2022) 'Time series analysis for temperature forecasting using TinyML', *Proceedings - IEEE Consumer Communications and Networking Conference, CCNC*. IEEE, pp. 691–694. doi: 10.1109/CCNC49033.2022.9700573.
- AMRC (2022) *i4.0 Dictionary*. Available at: https://www.amrc.co.uk/files/document/189/1522841973_i4.0_dictionary_DEV2.pdf.
- Andrade, P. *et al.* (2022) 'A TinyML Soft-Sensor Approach for Low-Cost Detection and Monitoring of Vehicular Emissions', *Sensors*, 22(10), pp. 1–18. doi: 10.3390/s22103838.
- Banbury, C. R. *et al.* (2020) 'Benchmarking TinyML Systems: Challenges and Direction'. Available at: <http://arxiv.org/abs/2003.04821>.
- Bauer, M., van Dinther, C. and Kiefer, D. (2020) 'Machine learning in SME: An empirical study on enablers and success factors', *26th Americas Conference on Information Systems, AMCIS 2020*, (August).
- Brown, P. (2019) *Wales 4.0: Delivering Economic Transformation for a Better Future of Work, Review of Digital Innovation for the Economy and the Future of Work in Wales*.
- Dat hone (2021) *THE CHALLENGES FACING MANUFACTURING IN 2022*, *Scientist*. Available at: <https://static1.squarespace.com/static/5da61091648a1b4992b8c73d/t/626684910ea7ec40f70731a3/1650885799167/White+Paper+-+Soft+Copy.pdf>.
- Delnevo, G. *et al.* (2021) 'Evaluating the practical limitations of TinyML: An experimental approach', *2021 IEEE Globecom Workshops, GC Wkshps 2021 - Proceedings*. IEEE. doi: 10.1109/GCWkshps52748.2021.9682101.
- Department of Education (2020) *Employer skills survey 2019*, *GOV.UK*. Available at: <https://www.gov.uk/government/collections/employer-skills-survey-2019>.
- Dwivedi, Y. K. *et al.* (2021) 'Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy', *International Journal of Information Management*, 57, p. 101994. doi: <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>.
- Edge Impulse (2022a) *Edge Impulse*. Available at: <https://www.edgeimpulse.com/>.
- Edge Impulse (2022b) *Imagine*. Available at: <https://edgeimpulse.com/imagine>.

FSB Wales (2003) *Federation of Small Businesses in Wales (2003) 'Manifesto for the National Assembly elections May 2003: A Winning Wales Complement*. Available at: <http://www.fsb.org.uk/policy/apu/devolved/wales/assets/MANIFenglish.pdf>.

Google Trends (2022) *TinyML - Google Trend*. Available at: <https://trends.google.com/trends/explore?date=today 5-y&q=TinyML>.

Grant Thornton UK LLP (2021) 'Establishing the labour availability issues of the UK Food and Drink Sector', (August), pp. 1–62. Available at: <https://commonslibrary.parliament.uk/research-briefings/cbp-8978/>.

Grau, M. M. and Centelles, R. P. (2021) 'TinyML: From Basic to Advanced Applications Author'.

Higginbotham, S. (2022) *TinyML needs a big use case. Or any use case., staceyoni*. Available at: <https://staceyoni.com/tinymml-needs-a-big-use-case-or-any-use-case/>.

IBM (2022) *What is Industry 4.0?* Available at: <https://www.ibm.com/uk-en/topics/industry-4-0>.

IBM Cloud Education (2020) *Neural Networks, IBM*. Available at: <https://www.ibm.com/uk-en/cloud/learn/neural-networks>.

Janapa Reddi, V. *et al.* (2022a) 'Widening Access to Applied Machine Learning with TinyML', *Harvard Data Science Review*. doi: 10.1162/99608f92.762d171a.

Janapa Reddi, V. *et al.* (2022b) 'Widening Access to Applied Machine Learning with TinyML', *Harvard Data Science Review*, 4(1), pp. 1–38. doi: 10.1162/99608f92.762d171a.

Jarrett, C. (2022) *Preventative Maintenance: From Downtime to Uptime, sme.org*. Available at: <https://www.sme.org/technologies/articles/2022/may/preventative-maintenance-from-downtime-to-uptime/>.

Kaiser, J. *et al.* (2021) 'Towards low-cost machine learning solutions for manufacturing SMEs', *AI and Society*. Springer London, (0123456789). doi: 10.1007/s00146-021-01332-8.

Khan, F. (2021) 'Start-Up To Scale-Up: Supporting Smes To Grow'.

MADE Cymru (2022) *ADOPTING DIGITAL TECHNOLOGY – A POLL WITH MANUFACTURERS IN WALES, Made Cymru*. Available at: <https://www.madecymru.co.uk/adopting-digital-technology-a-poll-with-manufacturers-in-wales/>.

Ng, A. (2022) *Machine Learning, Stanford University*. Available at: [http://mlclass.stanford.edu/#:~:text=What Is Machine Learning%3F,understanding of the human genome](http://mlclass.stanford.edu/#:~:text=What%20Is%20Machine%20Learning%3F,understanding%20of%20the%20human%20genome).

onrobot (2022) *FIVE BIGGEST CHALLENGES FOR SME'S IN 2022, onrobot.com*. Available at: <https://onrobot.com/en/blog/five-biggest-challenges-for-small-medium-size-manufacturers-in-2022>.

Ottesjö, B. *et al.* (2020) 'A tool for holistic assessment of digitalization capabilities in manufacturing SMEs', *Procedia CIRP*, 93, pp. 676–681. doi:

10.1016/j.procir.2020.03.078.

de Prado, M. *et al.* (2019) 'Bonseyes AI Pipeline -- bringing AI to you. End-to-end integration of data, algorithms and deployment tools', 1(4). doi: 10.1145/3403572.

Ray, P. P. (2022) 'A review on TinyML: State-of-the-art and prospects', *Journal of King Saud University - Computer and Information Sciences*. The Authors, 34(4), pp. 1595–1623. doi: 10.1016/j.jksuci.2021.11.019.

Ren, H., Anicic, D. and Runkler, T. A. (2021) 'TinyOL: TinyML with Online-Learning on Microcontrollers', *Proceedings of the International Joint Conference on Neural Networks*, 2021-July. doi: 10.1109/IJCNN52387.2021.9533927.

Richardson, M. (2022) *The data-driven course to net zero*, *aero-mag.com*. Available at: <https://www.aero-mag.com/the-data-driven-course-to-net-zero>.

Sawyer, D., Croome, M. and Jongboom, J. (2022) 'State of the tinyAutoML Market 2022 Edge Impulse / EON Tuner', pp. 1–20.

Schönfuß, B. *et al.* (2021) 'A catalogue of digital solution areas for prioritising the needs of manufacturing SMEs', *Computers in Industry*, 133. doi: 10.1016/j.compind.2021.103532.

Scopus (2022) *TinyML - Scopus Analysis*. Available at: <https://www.scopus.com/term/analyzer.uri?sid=73eca18aeca2d0159f91e2d9ed17a030&origin=resultslist&src=s&s=TITLE-ABS-KEY%28%22TinyML%22+OR+%22Embedded+Machine+Learning%22+OR+%22tiny+ml%22+OR+%22Tiny+Machine+Learning%22%29&sort=plf-f&sdt=b&sot=b&sl=94&count=>.

Shafique, M. *et al.* (2021) 'TinyML: Current Progress, Research Challenges, and Future Roadmap', *Proceedings - Design Automation Conference*. IEEE, 2021-Decem, pp. 1303–1306. doi: 10.1109/DAC18074.2021.9586232.

Shrivastav, S. (2020) *Artificial Intelligence, Machine Learning, and Deep Learning. What's the Real Difference?*, *Medium.com*. Available at: <https://medium.com/swlh/artificial-intelligence-machine-learning-and-deep-learning-whats-the-real-difference-94fe7e528097>.

Silva, M. C. *et al.* (2021) 'Wearable edge ai applications for ecological environments', *Sensors*, 21(15), pp. 1–31. doi: 10.3390/s21155082.

Sipola, T. *et al.* (2022) 'Artificial Intelligence in the IoT Era: A Review of Edge AI Hardware and Software', *Conference of Open Innovation Association, FRUCT*, 2022-April(i), pp. 320–331. doi: 10.23919/FRUCT54823.2022.9770931.

Stewart, M. (2020) *Tiny Machine Learning: The Next AI Revolution*, *towardsdatascience.com*. Available at: <https://towardsdatascience.com/tiny-machine-learning-the-next-ai-revolution-495c26463868>.

Su, W. *et al.* (2022) *AI on the edge: a comprehensive review*, *Artificial Intelligence Review*. Springer Netherlands. doi: 10.1007/s10462-022-10141-4.

Sudharsan, B. *et al.* (2021) 'TinyML Benchmark: Executing Fully Connected Neural Networks on Commodity Microcontrollers', *7th IEEE World Forum on Internet of Things, WF-IoT 2021*. IEEE, 0, pp. 883–884. doi: 10.1109/WF-IoT51360.2021.9595024.

Sudharsan, B. *et al.* (2022) 'OTA-TinyML: Over the Air Deployment of TinyML Models and Execution on IoT Devices', *IEEE Internet Computing*. IEEE, 26(3), pp. 69–78. doi: 10.1109/MIC.2021.3133552.

SurveyMonkey (2022) *SurveyMonkey*. Available at: <https://www.surveymonkey.co.uk/>.

Swansea University (2019) *ARTIFICIAL INTELLIGENCE: EMERGING CHALLENGES, OPPORTUNITIES, AND AGENDA FOR RESEARCH AND PRACTICE*. Available at: <https://www.swansea.ac.uk/research/research-highlights/digital-futures/artificial-intelligence/>.

SWMAS (2021) 'Manufacturing Barometer - Special Focus: Barriers to Growth Opportunities'. Available at: https://www.swmas.co.uk/sites/default/files/Q1_2021_Manufacturing_Barometer_-_National_Report.pdf.

Torres, Y., Nadeau, S. and Landau, K. (2021) 'Classification and quantification of human error in manufacturing: A case study in complex manual assembly', *Applied Sciences (Switzerland)*, 11(2), pp. 1–23. doi: 10.3390/app11020749.

Tyagi, N. (2020) *6 Major Branches of Artificial Intelligence (AI)*, *analyticssteps*. Available at: <https://www.analyticssteps.com/blogs/6-major-branches-artificial-intelligence-ai>.

Walters, A. T. (2008) 'The Impact of Advanced Manufacturing Technology on Small Welsh Companies', 1(May 2008).

Watney, C. and Auer, D. (2021) 'Encouraging AI adoption by EU SMEs', (March).

Welsh Government (2021) 'Manufacturing future for Wales : framework', (February).

Welsh Government (2022) *Welsh universities change lives with artificial intelligence and data science*, *gov.wales*. Available at: <https://gov.wales/welsh-universities-change-lives-artificial-intelligence-and-data-science>.

Wong, A. *et al.* (2020) 'TinySpeech: Attention Condensers for Deep Speech Recognition Neural Networks on Edge Devices', pp. 1–10. Available at: <http://arxiv.org/abs/2008.04245>.

Zimroz, R. and Krol, R. (2009) 'Failure analysis of belt conveyor systems for condition monitoring purposes', *Studia i Materiały*, (36), pp. 255–270.

Appendix A

The six branches of AI (Tyagi, 2020)

