LAB University of Applied Sciences

ARTIFICIAL INTELLIGENCE IN VISION-BASED APPLICATIONS

Tiivistelmä

Tekijä(t)	Julkaisun laji	Valmistumisaika
Lahtonen, Mino	Opinnäytetyö, AMK	Syksy 2020
	Sivumäärä	
	28	

Työn nimi

Tekoäly näköpohjaisissa sovelluksissa

Tutkinto

Tieto- ja viestintätekniikan insinööri (AMK)

Tiivistelmä

Opinnäytetyössä tutkittiin tämänhetkisen kehityksen huippua sovellutuksissa, jotka käyttävät näköön perustuvia sensoreita ja tekoälyä käsittelemään sensoreista saatua dataa. Toimeksiantajana opinnäytetyössä toimi LAB-ammattikorkeakoulun tekoälytiimi. Opinnäytetyössä käsiteltiin eri osa-alueiden taustoja, miten niissä hyödynnetään tekoälyä, mitä hyötyjä tekoälyllä saavutetaan ja miten tekoälyä voidaan tulevaisuudessa hyödyntää.

Opinnäytetyössä annettiin johdannon jälkeen lyhyt kartoitus tekoälyn historiaan, koneoppimiseen ja neuroverkkoihin. Tämän jälkeen käsiteltiin eri sovellutuksia eri osa-alueilla. Opinnäytetyön neljäs osuus käsitteli jo kehitteillä olevia tai konsepteja tulevaisuuden sovellutuksista. Opinnäytetyön loppuosassa tutkimustyön keskeiset kohdat vedettiin yhteen ja uusi tutkimuksen kohde ehdotettiin.

Opinnäytetyön tarkoituksena oli antaa tarpeeksi kattava katsaus kuvaa hyödyntäviin tekoälysovelluksiin, jotta sen pohjalta saataisiin tietoa ja ideoita, miten sitä voitaisiin käyttää ja viedä eteenpäin osa-alueille, joissa tekoälyä ei vielä täysin hyödynnetä.

Asiasanat

tekoäly, kamera, näkö, sovellus, neuroverkko, koneoppiminen, kuva

Abstract

Author(s)	Type of publication	Published
Lahtonen, Mino	Bachelor's thesis	Autumn 2020
	Number of pages	
	28	

Title of publication
Artificial intelligence in vision-based applications

Name of Degree
Information and Communications Technology

Abstract

The thesis researches the current state-of-the-art in applications that pairs vision-based sensors and artificial intelligence. The thesis is commissioned by the LAB University of Applied Sciences artificial intelligence team. The thesis covers the background of different fields, how they utilize artificial intelligence, what benefits does artificial intelligence bring and how can artificial intelligence be used in the future.

After the introduction of the thesis, the thesis gives a brief overview of artificial intelligence, machine learning and neural networks. After this, the thesis goes over applications in different fields. The 4th part of the thesis goes through future applications that are concepts or that field already being developed. In the conclusion of the thesis, the main parts of the thesis are summarized, and a new direction of research is suggested.

The purpose of the thesis was to provide a sufficiently comprehensive overview of vision-based artificial intelligence applications, to promote ideas to how to use artificial intelligence to fields where it is not yet fully utilized.

Keywords

Artificial intelligence, camera, vision, application, neural network, machine learning, image

CONTENTS

1 INTRODUCTION

The purpose of this thesis is to map the present state of the art in vision sensor based artificial intelligence applications and to explore possible future applications. The exponential growth in both hardware and software technology gives base for the use of artificial intelligence in virtually everything. Artificial intelligence is already being used to assist human workers, to make their job easier, to promote efficiency and precision and to make dangerous jobs less risky. However, all the benefits that artificial intelligence offers are not fully utilized yet. The usefulness of artificial intelligence can be pushed even further by offering it to different fields, that do not yet use artificial intelligence to its fullest potential. Optimizing the use of artificial intelligence frees humans to forget about their tedious jobs and focus on creative thinking and work, something that a machine cannot do.

The aim of the research is to find solutions and applications, that could help give ideas and information how different fields, that do not yet utilize artificial intelligence, would benefit from it. The goal is to give a base of knowledge, that could be then used to do more research or to develop or adapt a solution in some new field, using artificial intelligence. The information gathered in this thesis is from various digital sources, such as research articles, blog posts, case studies, state of the art papers and internet articles.

The questions for this research are: What is artificial intelligence? How is artificial intelligence implemented in different fields? What benefits does artificial intelligence bring? How can artificial intelligence benefit the future?

2   INTRODUCTION TO ARTIFICIAL INTELLIGENCE

2.1   History of artificial intelligence

Artificial intelligence means the simulation of human intelligence in computers that are modeled after humans, to think like them and to perform actions like them. The term "artificial intelligence" was born in 1956, when an American computer scientist John McCarthy suggested artificial intelligence to be removed from the field of cybernetics and given it its own field of research. (Ray, 2018.)

In the 1960s, artificial intelligence was already better at checkers than average human and it was used to solve mathematical problems and geometrical theorems. Computer scientists were also working on developing machine vision learning and applying it to robots. The research of artificial intelligence faced a problem with hardware, to research and create artificial intelligence, the applications needed to process enormous amounts of data, but the current computers at that era were not well-developed enough to do such job. This resulted in a shortage of funding that started from the mid-1970s and continued to mid-1990s and the research of artificial intelligence slowed down significantly. (Ray, 2018.)

Expert systems were created in the 1980s. They were used to simulate the knowledge and analytical skills of a human expert and their commercial success revived the research of artificial intelligence. Furthermore, the progress in CMOS transistor technology enabled the start of practical artificial neural network development. However, the collapse of Lisp Machine in 1987 market made artificial intelligence fall into a disrepute and made the research of artificial intelligence slow down once again. (Ray, 2018.)

The progress in computer hardware brought life to the research of artificial intelligence again in the late 1990s. Even with the dotcom bubble bursting in the early 2000s, machine learning continued its steady progress forward. Artificial intelligence was used in logistics, data mining and medical diagnosis and for example, the reigning world champion of chess in 1996, Garry Kasparov, was beaten by IBM's Deep blue (image 1). (Ray, 2018.)

Image 1. Garry Kasparov and IBM's Deep Blue (Ray 2018)

2.2   Machine learning

A branch of artificial intelligence, machine learning gives systems the ability to improve and learn from experience without being specifically programmed. Machine learning uses data it can access and develops computer programs based on what it learns from the data. The process begins by observing given data, such as instructions or examples, in order to look for patterns in data and then the application makes decisions based on the data. The purpose of machine learning is for computers to learn on their own and adjust their decision making, without the help of humans. (ExpertSystem, 2020.)

Machine learning functions are classified into many different categories, but there are three major categories, supervised learning, unsupervised learning and reinforcement learning. Supervised learning is when the algorithm builds mathematical models based on data that includes both the input and desired outputs, also called label. Supervised learning algorithms use the built model to predict future events or detect desired label from data it is fed (image 2). As an example, if the task is defining whether the image contains a cat, the training data would have images with or without a cat and each image would have a label if the image did include the cat or not. Semi-supervised learning has training data that is partially incomplete, portion of the data does not include the label. (Wikipedia, 2020.)
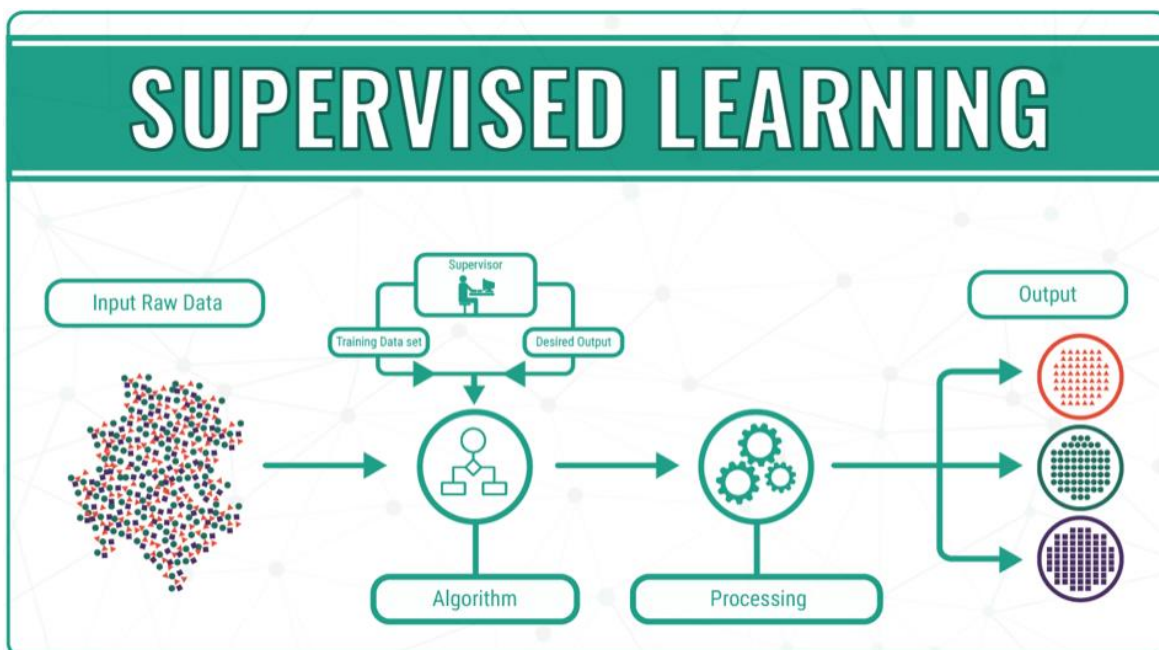
Image 2. Supervised learning diagram (Ghosh 2020)

Examples of supervised learning are classification algorithms and regression algorithms. Classification algorithm such as email filter have their output restricted to limited set of values. The input of an email filter would be an incoming email and the output would be the folder where to file the email. Regression algorithms in turn have repeated outputs, meaning they can have any value within a range. For example, temperature, length or a price of something are all continuous values. (Brownlee, 2016.)

Unsupervised learning is much like supervised learning, but its training data only includes inputs and no desired labels (image 3). This kind of algorithm is used to find hidden pattern in data, like grouping or clustering of data points. Cluster analysis identifies similarities in the data and acts based on the presence or absence of such similarities in every new bit of data. (Brownlee, 2016.)
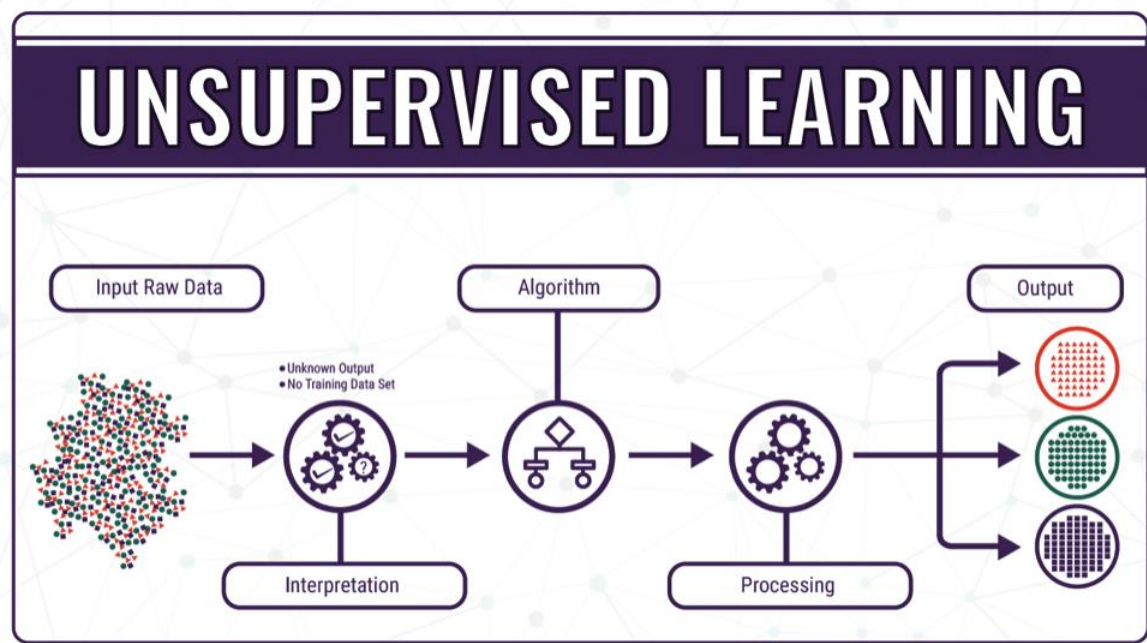
Image 3. Unsupervised learning diagram (Ghosh 2020)

Third major category of machine learning algorithm type is reinforcement learning. Its job is to take fitting action to maximize reward in a particular environment. It is used in various software and machines to find the optimal behavior or path it should take in a specific situation. The training data of reinforcement learning algorithm has no label, and the reinforcement agent decides what to do to perform the given task. Many reinforcement learning algorithms utilize dynamic programming techniques, and therefore the environment is typically state in a form of a Markov decision process (image 4). The mathematical framework is supported with the Markov decision process and it helps with decision making in situations where outcomes are under control of the decision maker or partly random. (Wikipedia, 2020.)
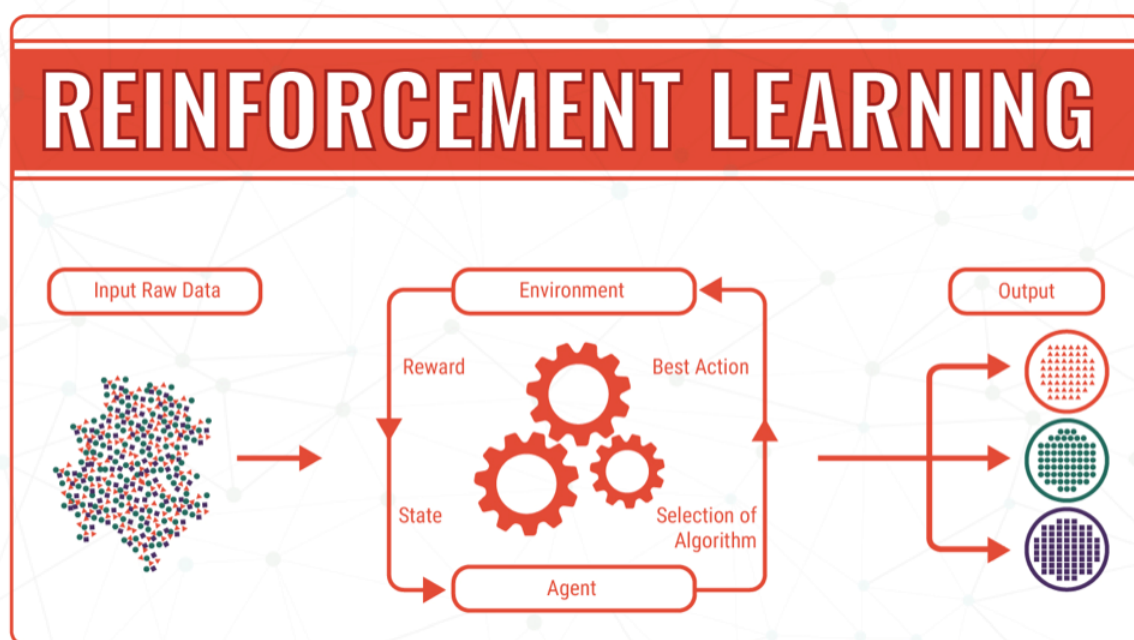
Image 4. Reinforcement learning diagram (Ghosh 2020)

Machine learning is widely used in industries that work with large amounts of data. Banks and other businesses in financial industries use machine learning to identify insights in data and to prevent fraud. In health care industry, machine learning is a growing fast thanks to wearable devices and sensor that can be used to determine the health of the user in real time. It is also used by medical experts to improve diagnosis and treatment and to reduce workload. Governments use machine learning to increase efficiency and save money and to detect fraud and identify theft. Websites give users tailored advertisement based on search or purchase history. Retailers use machine learning to personalize shopping experience, implement marketing campaigns, optimize prices, plan supply and to gain insight from customers. Data analysis in transportation is key to make routes more efficient and to predict potential problems to increase profits. (SAS, 2020.)

2.3 Deep learning

Deep learning is a machine learning algorithm that uses multiple layers to learn characteristics from data without the need of manual feature removal. While there are various deep learning models, vision based artificial intelligence applications use two the most: artificial neural networks and convolutional neural networks. (Wikipedia, 2020.)

2.3.1 Artificial neural networks

Artificial neural networks, also called connectionist systems, are brain-inspired (image 5) systems which are intended to replicate the way humans learn. They learn to perform

tasks by making errors and getting feedback, usually without being programmed with task-specific rule and then the system creates a model to use for future data that it compares to the created model. Neural networks consist of input and output layer and hidden layers consisting of units that transform the input for the output layer to use (image 5). (Medium, 2015.)
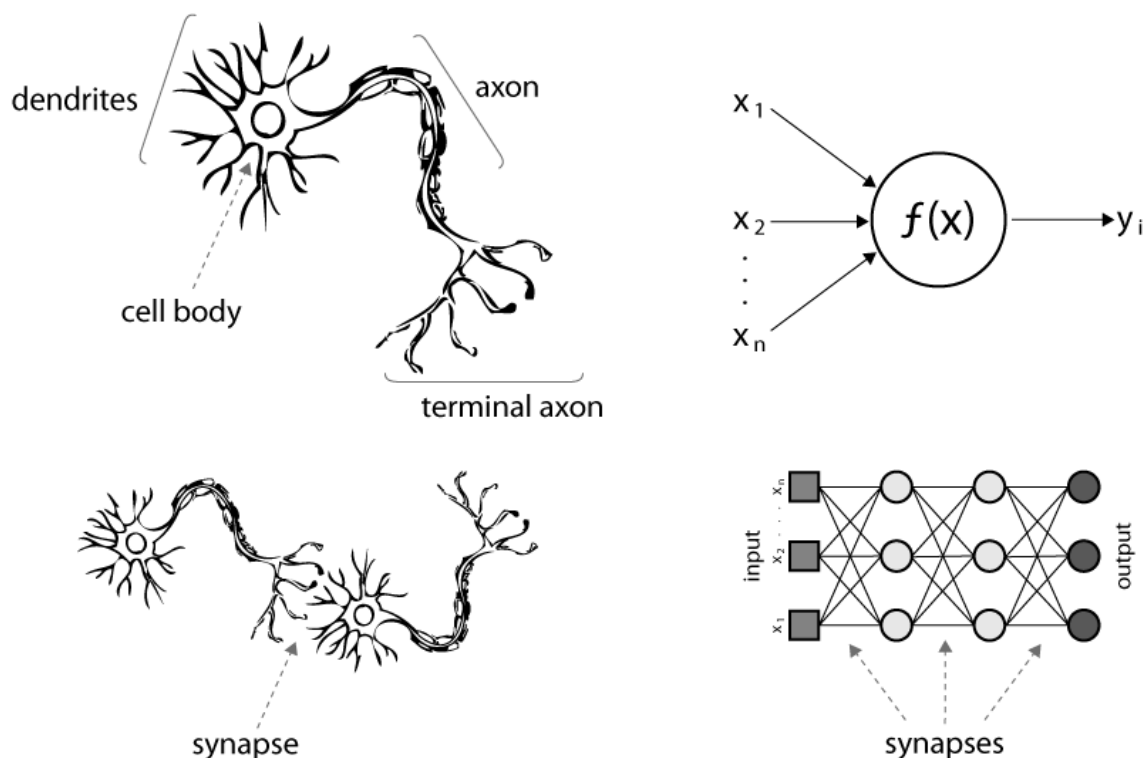

Image 5. Biological neurons and synapses versus artificial ones (Medium 2015)

Artificial neural networks have calculative units called neurons and those are connected with synapses, which are basically weighted values. When given a number, the neurons perform a calculation and then the result of this calculation is multiplied by a weight as it is put through the network. The result can be the output of your neural networks, but for more complex calculations more neurons and layers of neurons are needed. (Medium, 2015.)

Neural networks also use a technique called backpropagation, which allows the network to adjust their hidden layers of neurons if the outcome does not match the desired outcome. An artificial neural network is given a number of examples and then it is expected to get the same answer as the given example. When the answer is wrong, an error is calculated and the values at each neuron and synapse are propagated backwards through the network. For real world applications, the examples a process can take can be in millions. (Medium, 2015.)

2.3.2   Convolutional neural networks

Convolutional neural network has been the algorithm that has constructed and perfected computer vision with deep learning. The algorithm can take an image as an input and as- sign importance (learnable weights and biases) to various aspects or objects in the image and to be able to differentiate the image from another image. Normally the filters are hand-engineered, but with enough training, convolutional neural networks have the ability to learn these filters or characteristics by itself. By applying these filters, the algorithms can successfully capture spatial and temporal dependencies in an image. The reduced number of parameters and reusability of weights enable the architecture to perform a bet- ter fitting to the image dataset. Simply put, the network can be taught to understand the complexity of a picture better. (Saha, 2018.)

The design of convolutional neural networks is comparable to the connectivity pattern of neurons in the human brain and was modeled after the organization of the visual cortex. Stimuli elicit neuronal responses only in a fixed region in visual field known as receptive field. To cover the entire visual area, a collection of receptive fields overlaps each other. The process of convolutional neural network starts by giving the function an image that will then go through a series of convolution layers with filters. Then the results are pooled and fed to fully connected layers and then applying a softmax function, the image is given a probability value between 0 and 1 (image 6). (Saha, 2018.)
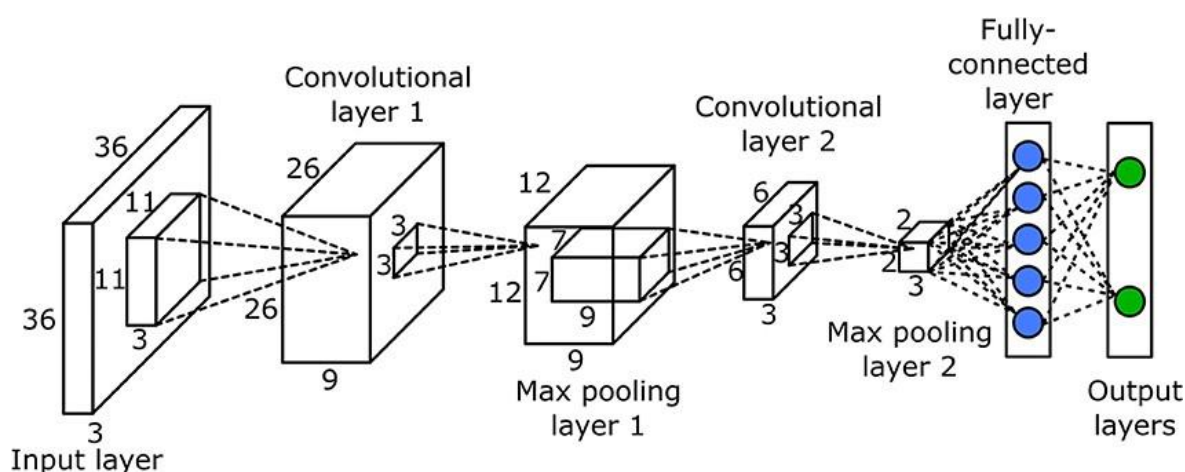

Image 6. Simple example of the architecture of a convolutional neural network. (Saha 2018)

Convolution layer extracts features from the input image and it preservers the relationship between pixels by learning features using small squares of input data. It takes two inputs

such as image matrix and a kernel or a filter. By applying filters, the convolution can perform operations such as blur, sharpen or edge detection. When a filter does not match the input image perfectly padding and valid padding is used. Padding adds zeroes next to pixels, so the filter fits and valid padding ignores the part of the image where the filter did not match. For non-linear operation, a rectified linear unit is used to introduce non-linearity to convolutional neural networks, since real world data wants the algorithm to learn would be non-negative linear values. (Saha, 2018.)

Pooling layers reduce the number of parameters when the images are too large. Spatial pooling, also called subsampling or downsampling, retains the important information of each map, but reduces the dimensionality. Three different types of spatial pooling are used: average pooling, max pooling and sum pooling. Average pooling calculates the average of the elements in a feature map, while max pooling takes the biggest element. Sum pooling sums all the elements in a feature map. (Saha, 2018.)

Fully connected layer flattens the feature map matrix into vector and feeds it into a fully connected layer like a neural network. The layer combines the features together and creates a model. Then softmax function is used to classify outputs as car, house, phone etc. (Saha, 2018.)

3 STATE OF THE ART IN VISION BASED ARTIFICIAL INTELLIGENCE APPLICATIONS

3.1 Healthcare

Current use of artificial intelligence in the medical field is still at a minimal level. Current applications are based on deep machine learning and image analyzing and are most used in diagnostics. (Jansen 2019, 33.)

3.1.1 Lunit

Lunit is a medical artificial intelligence software company founded in 2013, that focuses on cancer diagnosing. They develop and provide artificial intelligence powered solutions for cancer diagnostics and therapeutics. Lunit's INSIGHT CXR 3 solution uses deep learning technology to accurately detect 10 of the most common findings in a chest x-ray images. (Lunit Inc. 2019.)

The software creates heatmaps from the location information of detected lesions and gives an abnormality score that reflects the type of detected lesions. Then a case report is built that gives a summary of the analysis of each finding. This way the solution helps radiologists and clinicians in their interpretation process (Lunit Inc. 2019.)

The solution's main benefits lie in preventing more hard to notice cases of chest abnormalities from being missed while interpreting radiographs and increasing efficiency in interpretation through decreasing reading time by providing the user automatically generated case report. (Lunit Inc. 2019.)

3.1.2 PathAI

PathAI, company founded in 2016, develop their own convolutional neural network solution and computer vision algorithms to solve problems in diagnosing deadly diseases. In high-volume specimen cases, human pathologists need to examine close to 300 slides of specimen a day, each slide holding approximately 50,000 cells. Human mistakes cause the average error in diagnosis to rise to 15%. However, computers can work tirelessly and much faster in studying every pixel of an image and this results in PathAI's average diagnostics error rate to be 0.6%. The application is given images of the specimen and the artificial intelligence goes through the images and creates an overlay (image 7) of information on the image that the pathologist can then use to diagnose the case more accurately and faster. (Dougherty, 2019.)
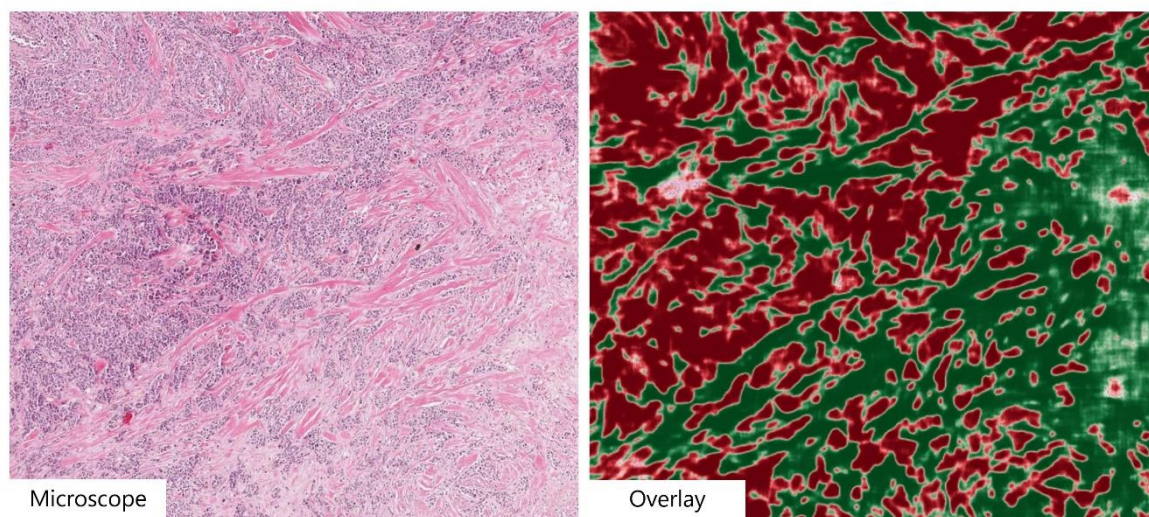
Image 7. Artificially generated overlay of an image (adapted Dougherty 2019)

PathAI's goal is to improve the speed and precision in high-volume specimen cases. This goal is achieved by offering the application for decision support and prognostic tests and use its help in identifying patients that could benefit from novel therapies, to make scalable personalized medicine a reality. PathAI is also extremely effective at assisting companies working with biomarkers to identify new drug pathways, while also improving patient outcomes and helping to improve the design for future clinical trials. (Dougherty, 2019.)

3.1.3 Face2Gene

Face2Gene offers a phenotyping mobile application that aids comprehensive and precise genetic evaluations. It detects phenotypes from facial images and automatically calculates anthropometric growth charts and matches the phenotype to genetic disorder based on gestalt. (Faggella, 2020.)

The application uses deep learning algorithms to build syndrome-specific computational-based classifiers (syndrome gestalts) and proprietary technology converts an image of the patient into de-identified mathematical facial descriptors. The facial descriptors are then compared to syndrome gestalts to quantify similarity, resulting in a prioritized list of syndromes with similar morphology. Artificial intelligence is also used to assist in feature annotation and syndrome prioritization by suggesting likely phenotypic traits and genes. (Face2Gene, 2020.)

3.2 Pose Estimation

Human pose estimation has been studied for over 15 years and remains one of the key problems in computer vision. It is very important for its abundance of applications that can

benefit from such technology. Most well-known application of pose estimation is motion capture technology but pose estimation can be also used in multitude of different applications. (Sigal, 2011.)

2-dimension pose estimation takes human joint information along the x and y axes, using computer vision. It can be used in applications that need behavior and gesture recognition or abstract human input into an environment. To fully capture the motion of a human, all three axes must be calculated (Babu, 2019.)

3-dimensional pose estimation must take into account a variety of factors and this is what makes 3-dimensional pose estimation such a challenging task and much harder than 2D pose estimation. Also, the lack of in the wild data is a major bottleneck. A 3D pose data set is built by using MOCAP systems which require elaborate setup with multiple sensors and bodysuit, and it makes them impractical to use outside. (Babu, 2019.)

3.2.1   Microsoft Kinect

Kinect, Microsoft's 3D pose estimation solution, brought the unique ability to reliably estimate the pose of a human user in real life, in any living room for entertainment purposes. The solution uses a RGBD camera and single frame tracking algorithm to achieve accurate and efficient 3D pose estimation. Efficiency was a key feature, since the solution was paired with Microsoft's gaming console Xbox 360, that has limited computational power. (Kohli & Shotton, 2013, 2.)

The system works by first estimating which body part each pixel belongs to and then using the information to estimate and calculate the location of each body joint using different machine learning methods (image 8). (Kohli & Shotton, 2013, 2.)
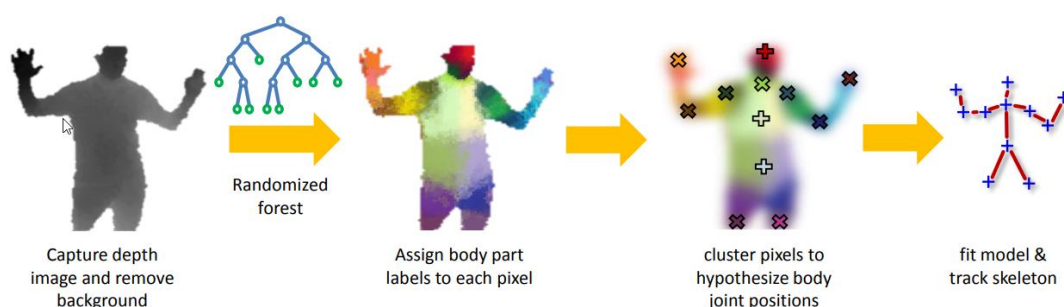

Image 8. Pipeline of the Kinect skeletal tracking system (Kohli 2013)

3.2.2 DeepMotion

DeepMotion, a company that launched a markerless AR avatar solution, is attacking a problem of generating 3D pose estimation from 2D video feed. Key challenge in generating joint depth from a 2D video is that the motions with non-lateral movements and limb occlusions makes it hard to create fluid 3D motions. (DeepMotion, 2019.)

DeepMotion's solution pairs vision model with character physics. Adding necessary body restrictions helps translate quick or complicated motions from 2D to 3D and with thoroughly simulated character, the results are limited to those that are physically possible. Their solution can be used for real-time user tracking, digital avatar creation or cheaper MOCAP solution for animations. (DeepMotion, 2019.)

3.3 Waste processing

230 million tons of trash is generated each year by consumers, businesses and the public sector in the USA. While there are specific efforts to reduce the amount of garbage produced over time, all communities will keep producing waste. So far, managing this waste has been largely a manual process, but some artificial intelligence applications have been utilized to automatize the process. The automatization has been done by including machine learning, deep learning and computer vision to the process of waste management. (Tractica, 2019.)

An example of modernizing waste process with artificial intelligence that does not include robotics, is a smart trash bin. A polish company, Bin.e, has developed an intelligent trash bin, that uses computer vision to identify the type of garbage thrown inside them. Their algorithm identifies and categorizes the type of thrash that is thrown away and then the waste is sorted into bins by type. This eliminates the need to sort through larger piles of waste at the waste processing centre. (Tractica, 2019.)

Further, the system senses when the bin is full, allowing for an optimized collection schedule. Collection route of the trash collection trucks can be optimized to only visit the locations where the bins are full, thus improving collection speed, lowering human labour costs and reducing fuel costs. (Tractica, 2019.)

3.4 Surveillance and Security

A.I is expected to revolutionize surveillance solutions. Systems based on artificial intelligence can analyze immense amounts of data and current camera technology makes that data much more accurate and diverse. Giving intelligence to the cameras is beneficial for

public safety, helping police and first responders more easily spot crimes and accidents, but it also raises questions about the future of privacy and social justice. (Jansen 2019, 36.)

3.4.1   SenseTime

SenseTime is currently one of the most valuable and biggest artificial intelligence companies. Based in China, the company has partners such as Massachusetts Institute of Technology (MIT), Qualcomm, Honda, Alibaba and Weibo. SenseTime has developed several artificial intelligence technologies including face, image, object and text recognition; medical image and video analysis; remote sensing; and autonomous driving systems. One of the biggest reasons SenseTime is so successful, is that they are government funded and have access to a database of 1.4 billion Chinese residents that they can use to teach their artificial intelligence solutions. (Marr, 2019.)

Several police departments in China use SenseTime's SenseTotem and SenseFace systems to analyse images and videos to catch offenders. SenseTotem uses a deep learning algorithm to perform content-based retrieval, identity verification and image search from database. To complement SenseTotem, SenseFace provides real time face recognition, facial capture, location tracking and data analysis for urban scenarios to provide a solution for public security, criminal investigation and city governance. (SenseTime, 2019.)

SenseTime also develops a computer vision analysis platform for city security management, called SenseFoundry. The platform is able to process more than 100 000 way of signals at a time and it can analyse over 100 billion unstructured and structured features. Combined with deep learning, SenseFoundry has functions like face recognition, vehicle recognition, passenger recognition, crowd analysis and event detection. The solutions real-time analysis of video streams frees human individuals who previously initiated these actions and lets them focus on more complex tasks and avoid thousands of hours of analysis. (EqualOcean, 2019.)

SenseVideo, a smart traffic monitoring solution, labels pedestrians and vehicles with small descriptions, allowing adminstrators to quickly scan through all video streams with these descriptions to locate suspects in hit-and-runs cases (image 9). SenseVideo can give over 10 distinctive attributes to pedestrians and vehicle attributes such as licence plate, vehicle model, vehicle color and vehicle brand. It also provides a reverse image search for pedestrians and vehicles. (SenseTime, 2019.)

Image 9. SenseVideo vehicle recognition system (SenseTime 2019)

3.4.2 BoulderAI

BoulderAI sells "vision as a service" solution. Their standalone proprietary camera has artificial intelligence built inside the camera and this eliminates the need of internet for the solution. They work with local government organizations to monitor and implement smart city solutions for roadways, venues and worksites among other locations. (BoulderAI. 2019.)

Single camera can monitor the capacity of parking lots, up to 60 spots at a time and a network of cameras allows the tracking of traffic in real time and over a period of time. With the help of artificial intelligence, the software can index vehicle occupancy, identify the model and make of the car and detect the age, mood and demographic of the passengers. (BoulderAI. 2019.)

3.4.3 IC Realtime

Established in 2006, IC Realtime is a digital surveillance manufacturer serving the residential, commercial, government and military security market. Through a partnership with a technology platform called Camio, IC Realtime created Ella. Ella is a cloud-based deep learning solution that enhances surveillance cameras with natural language search capabilities. (IC Realtime, 2020.)

The solution can recognize hundreds of thousands of natural language queries, letting users search footage to find clips showing specific animals, people wearing clothes of certain colour or individual car makes and models (image 10). (Vincent, 2019.)
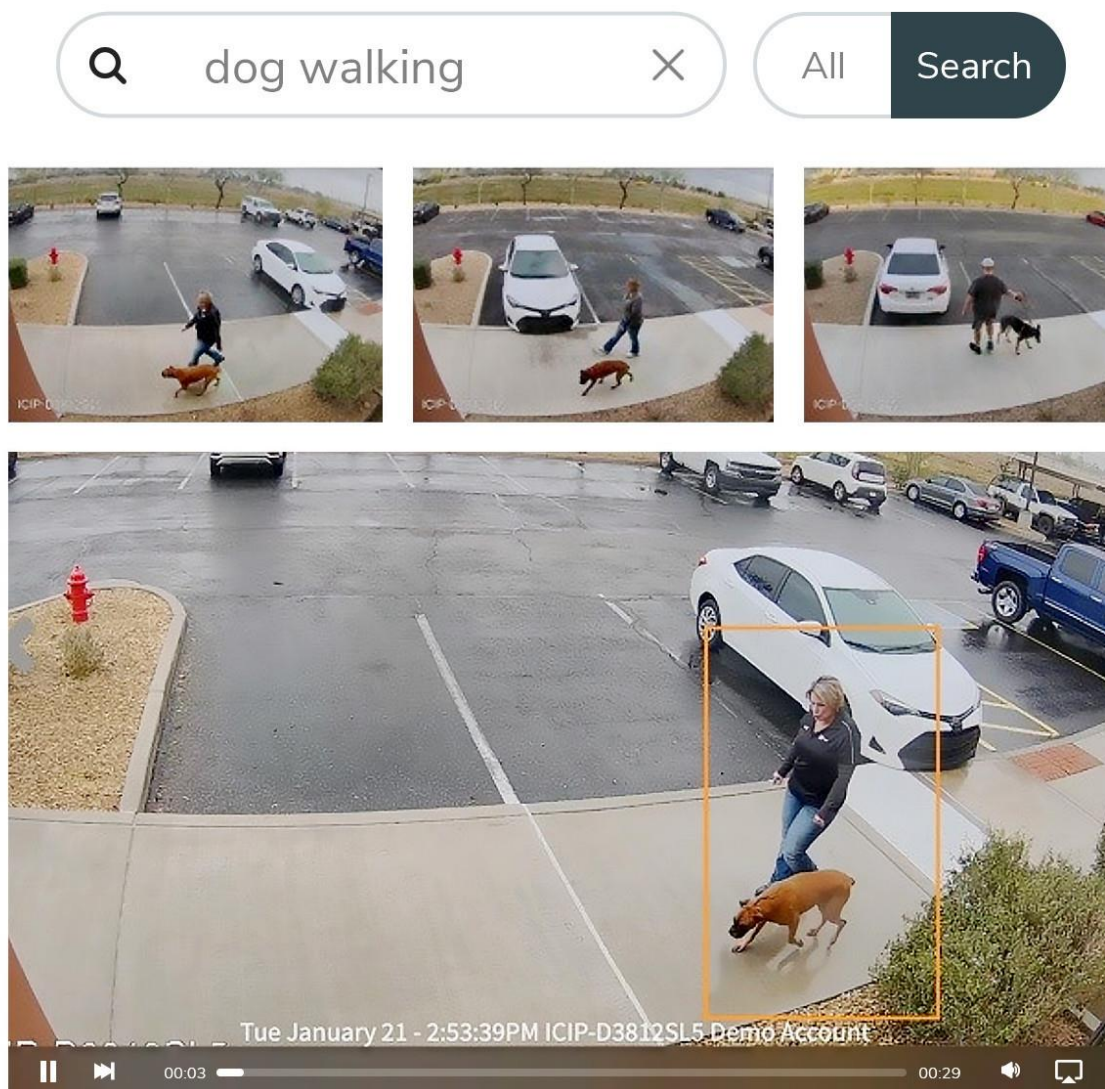
Image 10. Example of Ella's search function (IC Realtime 2020)

3.4.4 VALCRI

Visual analytics for sense-making in criminal intelligence analysis (VALCRI) is a semi-automated visual analytics system developed by the European Commission. VALCRI analyses from a wide range of mixed-format sources and displays its findings with easy-to-digest visualisations and comes up with possible explanations of crimes. Police agencies in many countries use it to speed up investigation, improve precision and pre-empt crimes

by detecting connections people often miss. One of the big benefits of VALCRI is that it also eliminates the bias and error of human effort. (VALCRI, 2020.)

The project was approached with cognitive engineering point of view. The idea of it is to bring the right resources for each of the tasks; humans for decision making and thinking, and machines for the repetitious manual labour, searching through millions of records of data in different databases and to present similar information within the same field. Concretely this mean splitting the workspace into three parts: the Data Space, the Analysis Space and the Hypothesis Space. The Data Space helps the analyst to see what data they have and how they are related, Analysis Space performs computations in order to understand relationships, trends and patterns and other significant behaviours and Hypothesis Space, where the analyst collects and assembles the data and formulates arguments and hypotheses that can be tested scientifically. Especially the Hypothesis Space is important, it brings the storytelling into crime and brings new perspectives and alternative points of view. (Cordis, 2018.)

Manually searching for related or relevant information in major cases requires an estimated 73 separate SQL queries and it can take up to five days. With VALCRI, this process can be performed in a blink of an eye. VALCRI's interaction design is based on tactile reasoning, to directly manipulate information in user interface. This helps the analysts to maintain and develop a summary of their investigative process whilst keeping track of the status of the process and identifying oversights and remaining tasks. Privacy protection is also one of VALCRI's main focuses, it automatically filters the faces the user can or cannot see in images and on video, it blurs the faces of those specified. (Cordis, 2018.)

3.4.5 Onetrack.AI

Onetrack.AI is developing a system that aims to remove forklift accidents. With machine learning and low-cost cameras, the system detects, records and prevents workplace accidents associated with forklifts. Their proprietary algorithm factors in safety, productivity and skill-levels of every employee to find opportunities to improve safety and efficiency. (OneTrack, 2019.)

The system sees which way the operator is looking, how fast the operator is going, where the operator is driving, people around the operator and how other forklift operators are manoeuvring their vehicles. It also detects safety violations, such as cell phone use and notify warehouse managers so they can take immediate action. (OneTrack, 2019.)

3.5 Agriculture

Much of the use of artificial intelligence in agriculture is focused on applications in robotics. Outside robotics, artificial intelligence is used in crop and soil health monitoring and predictive analysis. Deep learning systems can identify possible defects and nutrient deficiencies in soil through a camera of a user's smartphone or from satellite imagery. Predictive analysis is done with the help of satellite data to predict weather, analyse crop sustainability and evaluate farmlands for the presence of diseases and pests. (Jansen 2019, 42.)

3.5.1 Plantix

Berlin-based agricultural tech startup PEAT has developed a deep learning application called Plantix that identifies potential defects and nutrient deficiencies in soil. The analysis is done by software algorithms which correlate foliage patterns with certain soil defects, plant pests and diseases. (Fagella, 2020.)

The application identifies possible defects through the camera of a smartphone and provides the users soil restoration techniques, tips and other possible solutions. Plantix also aids farmers by recommending targeted biochemical or chemical treatments, reducing the number of agrochemicals in groundwaters and waterways that can result from overuse or incorrect use of herbicides and pesticides. (Tibbetts, 2018.)

3.5.2 Vineview

With NASA based technology, aerial-based sensors, ultra-high-resolution imagery and advanced scientific algorithms, Vineview brings a new perspective to the vineyard operators monitor grapevine health. Vineview provides the user scientifically calibrated vine vigour maps that gives them information to improve grape quality and yield, conserve water and reduce costs associated with harvest segmentation, homogenization, canopy management, fertilizer optimization and irrigation scheduling (image 11). (VineView, 2020.)

The application also helps prevent the spread of disease with early identification of symptomatic vines, georeferenced for targeted follow up, testing and removal. The technology is more efficient, accurate and reliable than highly trained experts scouting on the ground. (VineView, 2020.)
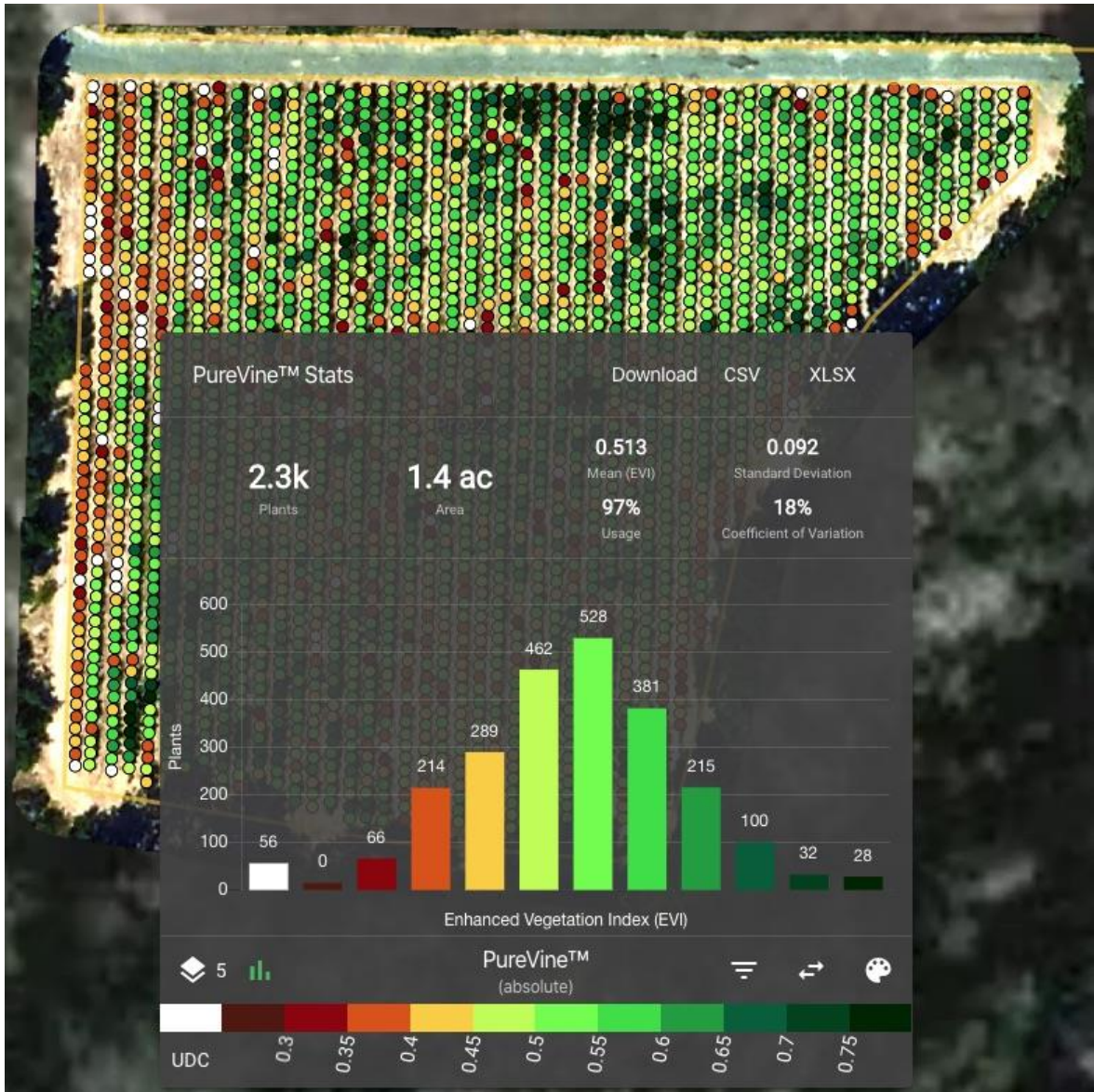
Image 11. Productivity and field efficiency analysis made by Vineview (Vineview 2020)

4   FUTURE APPLICATIONS

4.1   Automated medical image interpretation

While already used in some applications, automated medical image analysis has huge potential to help the currently overworked medical personnel in the future. Images are the biggest source of data and the hardest to analyze. Analyzing can take hours of valuable labor that could be instead used on helping patients. (Kuflinksi, 2018.)

Automated image analysis can help the strenuous work of radiologists that scrutinize over every image in search of anomalies and take off a part of their already out of control growing workload. Artificial intelligence could even flag the type of detected abnormalities such as cancer tumor being malignant or benign. This would further ease doctor's analytic workload and support them in decision making and help them focus more on patients (Kuflinksi, 2018.)

Artificial intelligence in image analysis software will change the roles of radiologists and other clinicians. Radiologists can focus on diagnosis and decision making instead of spending time screening images. The same technology will augment other clinicians with digital assistance to more easily read medical images, eliminating the need for radiologists to analyze the image and then give their information back to the clinicians. Artificial intelligence would enable all doctors and even paramedics, to be able to understand images from ultrasound scanners. (Kuflinksi, 2018.)

With artificial intelligence, patients will get faster results from x-ray's and will be presented timelier and more accurate diagnoses in the future. In hospital environment, patients will need fewer invasive procedures and the amount of radiation from x-rays are reduced, since less scans are required. (Kuflinksi, 2018.)

4.1.1   Aidoc

Aidoc develops artificial intelligence-based decision support software. Their technology analyses medical imaging to provide comprehensive solutions to help radiologists prioritize critical and urgent cases where faster diagnosis and treatment could save lives (image 12). This means that the software is going through the images on the background and detects abnormalities, even before the radiologist has access to the images. In addition to providing support for CT scans, support for oncology, X-Ray and MRI are being developed. The oncology solution will instantly detect, measure and compare tumour size with past scans. What makes Aidoc special, is the ability to analyse entire 3D CT scans,

instead of just 2D slices. This means that radiologists no longer need to switch between discrete image analysis applications. (Kuflinksi, 2018.)
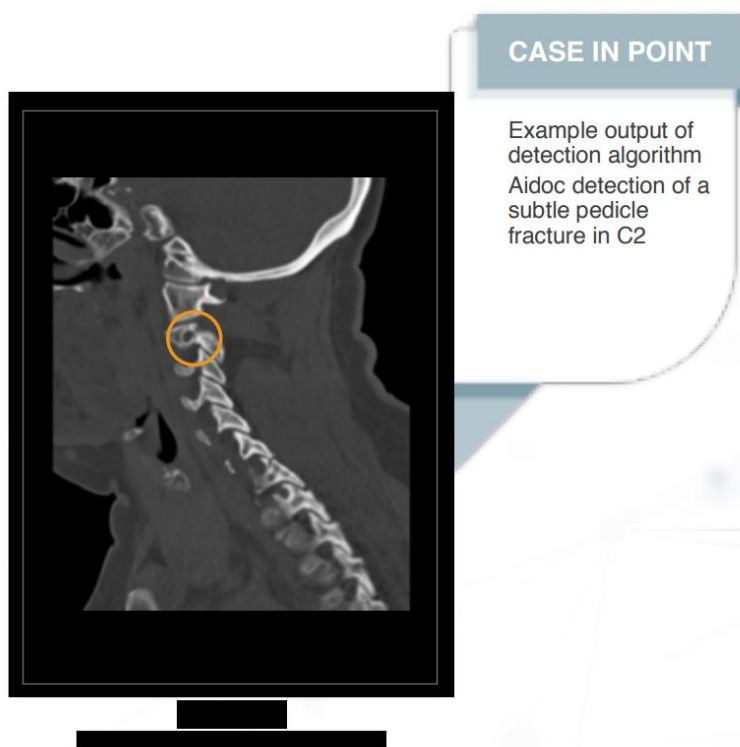

Image 12. Example output of detection algorithm (Aidoc 2019)

4.1.2 ViosWorks

Historically, acquiring quality cardiac sequences to diagnose cardiac function and flow has been a difficult and time-consuming exam to perform. Current cardiac MRI techniques require multiple slice acquisitions that are perpendicular to the flow of blood. This means that the patient is required to hold their breath in some pathologies, and multiple times an exam. For patients already with heart disease, this is very difficult, and the exams suffer from below average or non-diagnostic image quality. (Umairi, 2019.)

Combating this problem with machine learning, San Francisco company Arterys is developing a solution that aims to eliminate some of the current complicating circumstances that makes cardiovascular magnetic resonance imaging lengthy and difficult. A beneficial by-product of their solution is that it also provides increased quantity and quality of the data, compared to the conventional methods (Kuflinksi, 2018). The Royal Hospital in Muscat, Oman, has implemented Arterys' ViosWork 4D Flow sequence into their standard

cardiovascular magnetic resonance exams. With this technique, the data acquisition is made completely without breath-holds and there is little interaction necessary on the front end and the images are immediately reconstructed for inspection and analysis. The volume of data from ViosWork 4D flow is enough to cover the entire chest and does not require separate acquisitions from different locations like traditional 2D cardiovascular magnetic resonance sequences does (image 13). (Umairi, 2019.)
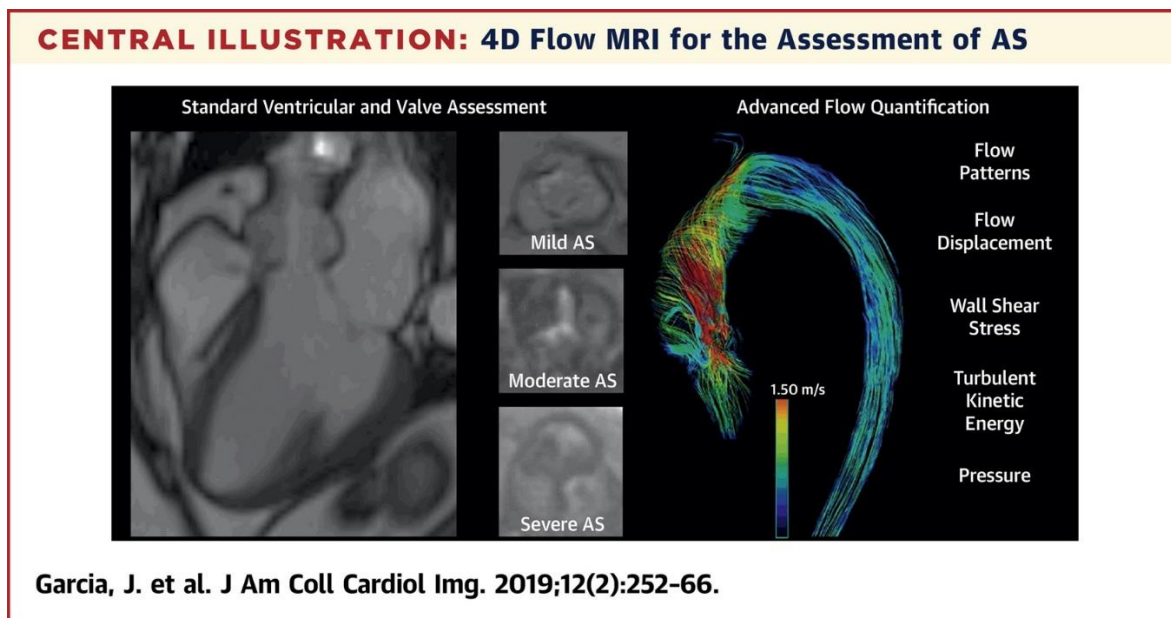

Image 13. Comparison of standard assessment and advanced flow quantification (Garcia 2019)

4.1.3   Butterfly iQ

Butterfly Network is developing an ultrasound probe called iQ. Normally, ultrasound images are generated by using an oscillating piezoelectric crystal that produces waves. Instead of piezoelectric crystals, iQ is using semiconductor chips to produce the sound waves. Due to the efficiency of semiconductor fabrication, the hardware iQ uses makes it more than 50 times cheaper than traditional ultrasound units. The semiconductor approach also means that the iQ follows Moore's law and the hardware will get exponentially cheaper and better with time. (Seals, 2017.)

Another benefit that using semiconductors has, is that it offers a wider acoustic bandwidth and that eliminates the need for multiple probes. Using traditional ultrasound means that you need to use variety of probes for different exams, due to the narrow bandwidth that the devices have. For example, one probe to get the waves deep into the abdomen or one probe for superficial thyroid gland. (Seals, 2017.)

On top of the advanced hardware, iQ uses deep learning to analyse the ultrasound images that the probe provides (image 14). With the help of artificial intelligence, iQ gives the ability to clinicians to figuratively "look" inside the patient. Ultrasound imaging augmented with artificial intelligence could possibly replace the stethoscope as the fundamental tool of all clinicians, meaning more efficient and precise diagnoses. Eliminating the need for ultrasound training to diagnose the ultrasound imagery would also mean that it could bring ultrasound devices to every home. (Seals, 2017.)

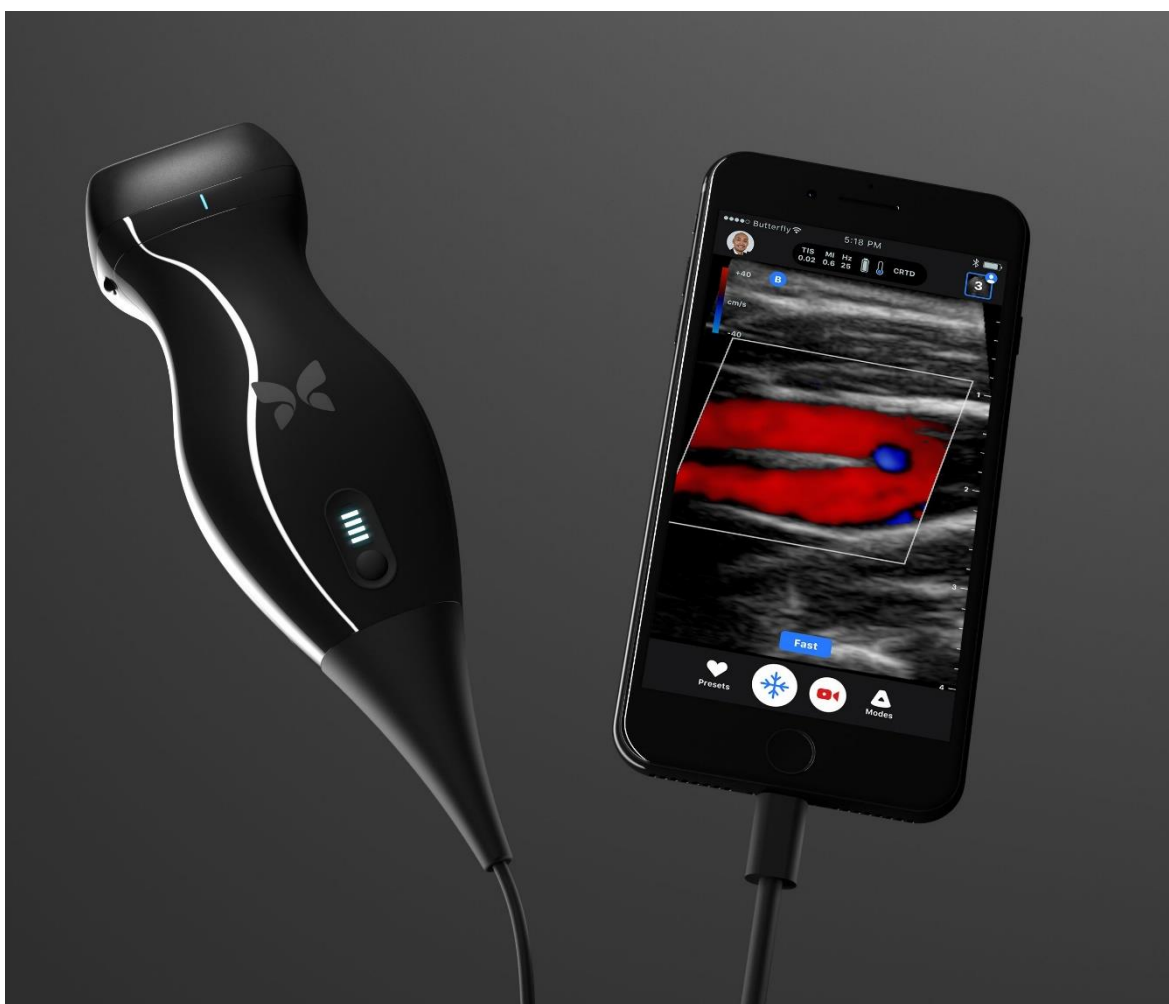Image 14. Butterfly Network iQ and artificial intelligence made overlay on ultrasound imagery (CapeRay 2017)

4.2   Urban decay and growth

Massachusetts Institute of Technology's Media Lab developed a computer vision system 7 years ago, that analyses images of streets in urban neighbourhoods in order to scope how safe the neighbourhoods appear. Together with Harvard University, the MIT team

further used this system to identify factors that predict urban change. They used 1.6 million pairs of photos taken seven years apart, to determine the deterioration or improvement of neighbourhoods in different American cities. The results were then used to test different hypotheses in social sciences about the cause of urban restoration. The results showed that contrary to popular belief, raw income levels or housing prices of the neighbourhoods do not predict change, but the amount of highly educated residents and the closness to central business districts correlate strongly with improvements in physical condition. (Hardesty, 2017.)

Previously, the system was trained with hundreds of thousands of examples of safe streets, manually rated by human volunteers. In the new research, the system compared the same pictures with pictures from Google's Street View visualisation tool, from same geographic coordinates and seven years apart. The images were pre-processed, because the system might give false results based on if one of the pictures in the pairs would be taken in winter or summer. This is due to the system's previous training to flag green spaces as one of the criteria of assessing safety and the difference in season might make the system think, that the place had lost green space and give unreliable results. (Hardesty, 2017.)

A technique called semantic segmentation was used to classify every pixel in the pictures to reject pictures that had some object obscuring majority of the picture. The system would then compare some other picture from different coordinates on the same block. To validate the systems results, 15,000 randomly selected pairs of images were selected and presented to voluntary reviewers. The volunteers were asked to determine the relative safety of the neighbourhoods depicted. Their review corresponded with the system's 72 percent of the time. (Hardesty, 2017.)

4.3   Instant automatic photo editing

For years it has been assumed that the replacement for Photoshop would be another new app for Windows or Mac computer. But the replacement is not that distant into the future and it is something already demonstrated on smart phones and their cameras, with the help of artificial intelligence and machine learning. The camera software on smartphones is already able to automatically correct an image before even showing it to the user. Much of the retouching process done on Photoshop can be replaced by real-time computational process which improves the taken picture in less time than it takes for the camera shutter to open and close. (Smith, 2017.)

What makes the artificial solution better compared to manual editing on Photoshop, is that the real-time corrections smart phones can make are applied at the pixel level, not globally. An edit on Photoshop typically impacts the entire image, while an algorithm showcased by Google and MIT researchers can apply the edit on specific areas, only where they are needed. This kind of algorithm was created by feeding a neural network 5000 images that had been professionally edited. This taught the neural network the attributes of a high-quality image. Additionally, machine learning could be used to teach an image-editing neural network to understand different editing styles of photographers and photo editors. (Smith, 2017.)

The MIT and Google mentioned project is to develop an app to analyse images to determine which parts need retouching. The algorithm looks for the same areas that a Photoshop professional would look for, colour, saturation and lightness. It then uses the rules from the neural network to make changes that are then applied to the resulting picture. (Smith, 2017.)

4.4   Emotion artificial intelligence

Emotions and cognitive functions form the human intelligence and emotions play a significant role in human communications and cognition. People react to various feelings and internal or external stimuli by showing emotion and this makes human a very special compared to other living beings. Officially, emotion is a biological state associated with nervous system, established by neurophysiological changes associated with behavioural responses, thoughts and feelings. Research on emotions has increased significantly in the fields like psychology, neuroscience and computer science. Using artificial intelligence, researchers aim to develop systems that can recognise, interpret, process and simulate human emotions. The goal is for the system to understand feelings when they see it and react appropriately.

Emotion artificial intelligence has many potential benefits to the real world. There are already some virtual tutors that can be used to help the process of learning. But with the help of artificial intelligence, the virtual tutors could be augmented with eyes, to see when the students are bored, frustrated or confused. This way the virtual tutor could tailor the learning experience better and encourage the student not to give up. In business, customers emotion recognition to products and product ads can have significant impact in improving their services and marketing. Analysing more hidden emotions like depression and anxiety, can be very beneficial to people's emotional wellbeing and assessment of mental problems.

Emotion artificial intelligence works by using supervised learning. The system is given a group of predefined discrete emotions and then a computational model is built to predict the emotions. One's emotional expression can be determined based on factors such as location of eyebrows, eyes and the movement of mouth. The system learns to determine the association between an emotion and its external demonstration from large collection of labelled data. When the data is compared to provided labels, the computational model will adjust the parameters based on the predicted error.

Humans are very good at hiding their emotions. Therefore, it is extremely difficult for an artificial intelligence to truly understand human emotion just from face alone. Micro-expressions are involuntary facial expressions that can reveal suppressed emotional states, but adding other cues, such as voice, gestures, action and heart rate, it becomes much easier for the computational model to recognise and understand human emotion. With more clues, a more reliable system can be built, to really understand human emotions.

5   CONCLUSION

The purpose of the thesis was to map the current state-of-the-art of vision-based applications that use artificial intelligence. The research gives a brief introduction to artificial intelligence, machine learning and neural networks and shows how these areas of computer science are utilised in different fields and what benefits they bring to everyday life. Searching for information, to be used in the thesis, proved to be difficult, because most of the websites introducing the applications were very commercialised and required strict analysis and filtering, what could be used and what not.

The benefits of artificial intelligence are significant. Machine vision paired with artificial intelligence is faster, it is more precise, and it is more efficient than human eyes and labour. Machine vision gives humans perspectives that we cannot see ourselves and those perspectives help us make decisions faster, better and pre-emptively. Using artificial intelligence in surveillance brings more security and helps to prevent accidents and crimes. Utilising artificial intelligence to do work that requires manual, repetitious actions, human workers are freed from the tedious and boring work and they can focus on tasks that require more creative thinking, tasks that machines cannot do. This makes workplace atmosphere better and promotes safety.

The use of artificial intelligence is especially useful in the field of healthcare. Using artificial intelligence, doctors are potentially freed from the most burdening part of their work, diagnosing and analysing samples. Trained artificial intelligences can go through and analyse significantly higher number of samples and doing so with smaller margin of error. This frees doctors to focus on helping and saving the patients. Artificial intelligence is also revolutionising the data that comes from different medical examination machines. Artificial intelligence is able to analyse the data and give doctors specific samples that it has flagged for inspection and it can create an overlay on that sample to further pinpoint the point of interest, to lessen the extent of analysis by the doctor.

The advances in technology will further the development of artificial intelligence. Because of Moore's law, computer hardware continues to exponentially get cheaper and more powerful and when artificial intelligence systems get access to more powerful hardware, training and building them gets easier, cheaper and more advanced systems can be built to tackle more difficult problems.

The information given in this thesis can create ideas how to use artificial intelligence in some fields where it is not fully utilised yet. Or how some of the mentioned applications could be altered to serve some other function in some other field. To further research this

subject, a comprehensive study on neural networks and machine learning with machine vision could be conducted and that research could be used to apply neural networks and deep learning into fields and applications that do not yet use them.

REFERENCES

Digital sources

Aidoc. 2019. Next Gen Radiology AI. Whitepaper.

Babu, C. S. 2019. A 2019 guide to 3D Human Pose Estimation. Blog. Nanonets. Referenced 16.1.2020. Available at https://nanonets.com/blog/human-pose-estimation-3d-guide/

Babu, C. S. 2019. A 2019 guide to Human Pose Estimation with Deep Learning. Blog. Nanonets. Referenced 16.1.2020. Available at https://nanonets.com/blog/human-pose-estimation-2d-guide/

BoulderAI. 2019. Solutions. BoulderAI. Referenced 17.1.2020. Available at https://www.boulderai.com/solutions/

Brownlee, J. 16.3.2016. Supervised and Unsupervised Machine Learning Algorithms. Blog. Machine Learning Mastery. Referenced 10.5.2020. Available at https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

CapeRay. 1.12.2017. Ultrasound on a Chip. Blog. Referenced 25.11.2020. Available at https://www.caperay.com/blog/index.php/2017/ultrasound-on-a-chip/

Cordis. 30.6.2018. Visual Analytics for brighter criminal intelligence. Cordis. Referenced 20.1.2020. Available at https://cordis.europa.eu/article/id/218541-visual-analytics-for-brighter-criminal-intelligence

DeepMotion. 21.5.2019. Using Markerless Augmented Reality for Digital Avatars. DeepMotion. Referenced 17.1.2020. Available at https://blog.deepmotion.com/2019/05/21/markerless-augmented-reality-for-ar-avatars/

Dougherty, E. 12.11.2018. Artificial intelligence decodes cancer pathology images. Referenced 15.1.2020. Available at https://www.novartis.com/stories/discovery/artificial-intelligence-decodes-cancer-pathology-images

EqualOcean. 2019. China's AI Giant SenseTime launches 11 New AI Products. EqualOcean. Referenced 18.1.2020. Available at https://equalocean.com/ai/20190515-chinas-ai-giant-sensetime-launches-11-new-ai-products

ExpertSystem. 2020. What is Machine Learning? A definition. Blog. ExpertSystem. Referenced 10.5.2020. Available at https://expertsystem.com/machine-learning-definition/

Face2Gene. 2020. Technology. Referenced 15.1.2020. Saatavilla
https://www.face2gene.com/technology-facial-recognition-feature-detection-phenotype-analysis/

Faggella, D. 18.5.2020. AI in Agriculture – Present Applications and Impact. Emerj. Referenced 5.6.2020. Available at https://emerj.com/ai-sector-overviews/ai-agriculture-present-applications-impact/

Faggella, D. 14.3.2020. Machine Learning for Medical Diagnostics – 4 Current Applications. Referenced 16.1.2020. Available at https://emerj.com/ai-sector-overviews/machine-learning-medical-diagnostics-4-current-applications/

Garcia, J., Barker, J., A., Markl, M. 2019. The Role of Imaging of Flow Patterns by 4D Flow MRI in Aortic Stenosis. State-of-the-art paper. ScienceDirect. Referenced 24.11.2020. Available at https://www.sciencedirect.com/science/article/pii/S1936878X1831115X

Ghosh, K. 12.7.2020. Introduction to Machine Learning. Medium. Referenced 26.11.2020. Available at https://medium.com/@ghoshkoustav18/introduction-to-machine-learning-c15c979709c8

Grieve, G. 17.7.2020. Explained SIEMply: Machine Learning. Blog. LogPoint. Referenced 26.11.2020. Available at https://www.logpoint.com/en/blog/explained-siemply-machine-learning/

Hardesty, L. 6.7.2017. Why do some neighborhoods improve? MIT News. Referenced 26.11.2020. Available at http://news.mit.edu/2017/highly-educated-residents-neighborhoods-improve-0706

IC Realtime. 2020. IC Realtime. Referenced 20.1.2020. Available at
https://icrealtime.com/

Jansen, P. 13.4.2018. D4.1: State-of-the-art Review. Sienna. Referenced 15.1.2020. Available at https://www.sienna-project.eu/digitalAssets/787/c_787382-l_1-k_sienna-d4.1-state-of-the-art-review--final-v.04-.pdf

Kohli, P., Shotton J. 2013. Key Developments in Human Pose Estimation for Kinect. Microsoft. Referenced 17.1.2020. Available at https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/ks_book_2012.pdf

Kuflinski, Y. 8.11.2020. How Medical Image Analysis Will Benefit Patients and Physicians. Blog. Iflexion. Referenced 3.6.2020. Available at https://www.iflexion.com/blog/medical-image-analysis

Lunit Inc. 2019. AI Product Insight CXR3. Referenced 15.1.2020. Available at
https://www.lunit.io/en/product/insight_cxr3/

Marr, B. 17.6.2019. Meet The World's Most Valuable AI Startup: China's SenseTime.
Blog. Forbes. Referenced 18.1.2020. Available at https://www.forbes.com/sites/bernard-
marr/2019/06/17/meet-the-worlds-most-valuable-ai-startup-chinas-
sensetime/#2ba70a26309f

Medium. 28.12.2015. Everything You Need to Know About Artificial Neural Networks. Me-
dium. Referenced 19.5.2020. Available at https://medium.com/technology-invention-and-
more/everything-you-need-to-know-about-artificial-neural-networks-57fac18245a1

OneTrack. 5.6.2019. How Computer Vision and Deep Learning improve Forklift Safety.
OneTrack. Referenced 5.2.2020. Available at https://www.onetrack.ai/post/how-computer-
vision-and-deep-learning-improve-forklift-safety

Peng, W., Zhao, G. 19.11.2020. Artificial intelligence understands emotions. Blog. Univer-
sity of Oulu. Referenced 27.11.2020. Available at https://www.oulu.fi/blogs/science-with-
arctic-attitude/emotion-ai

Ray, S. 11.8.2018. History of AI. towards data science. Referenced 10.5.2020. Available
at https://towardsdatascience.com/history-of-ai-484a86fc16ef

Saha, S. 15.12.2018. A Comprehensive Guide to Convolutional Neural Networks – the
ELI5 way. Medium. Referenced 19.5.2020. Available at https://towardsdatascience.com/a-
comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

SAS. 2020 Machine Learning, What it is and why it matters. SAS. Referenced 12.5.2020.
Available at https://www.sas.com/en_us/insights/analytics/machine-learning.html

Seals, K. 6.12.2017. Ultrasound-on-a-chip supercharged with AI: The most disruptive
technology in radiology?. Becoming Human. Referenced 25.11.2020. Available at
https://becominghuman.ai/ultrasound-on-a-chip-supercharged-with-ai-the-most-disruptive-
technology-in-radiology-b2684b0421aa

SenseTime. 2019. SenseTotem. SenseTime. Referenced 18.1.2020. Available at
https://www.sensetime.com/en/Service/Security_SenseTotem.html#product

SenseTime. 2019. SenseVideo. SenseTime. Referenced 18.1.2020. Available at
https://www.sensetime.com/en/Service/Security_SenseVideo.html#product

Smith, J. 14.8.2017. Future of Photoshop - replaced by AI and Machine Learning. Ameri-
can Graphics Institute. Referenced 26.11.2020. Available at

https://www.agitraining.com/adobe/photoshop/classes/future-of-photoshop-replaced-by-ai-machine-learning#mainContent

Sigal, L. 2011. Human pose estimation. Disney Research. Referenced 16.1.2020. Available at https://www.cs.ubc.ca/~lsigal/Publications/SigalEncyclopediaCVdraft.pdf

Tibbetts, J., H. 12.1.2018. From identifying plant pests to picking fruit, AI is reinventing how farmers produce your food. Eco-Business. Referenced 18.5.2020. Available at https://www.eco-business.com/news/from-identifying-plant-pests-to-picking-fruit-ai-is-reinventing-how-farmers-produce-your-food/

Tractica. 2019. Modernizing Waste Management Through AI. Tractica. Referenced 17.1.2020. Available at https://www.tractica.com/artificial-intelligence/modernizing-waste-management-through-ai/

Umairi, A., R. 2019. A 10-minute comprehensive cardiac MR exam with flow quantification. Case study. GE Healthcare. Referenced 24.11.2020. Available at https://www.gesignapulse.com/signapulse/autumn_2019/MobilePagedArticle.action?articleId=1541689#articleId1541689

VALCRI. 2020. VALCRI. Referenced 20.1.2020. Available at http://valcri.org/about-valcri/

Vincent, J. 23.1.2018. Artificial Intelligence is Going to Supercharge Surveillance. Referenced 19.1.2020. Available at https://www.theverge.com/2018/1/23/16907238/artificial-intelligence-surveillance-cameras-security

VineView. 2020. Vine Vigor Products. VineView. Referenced 18.5.2020. Available at https://www.vineview.com/data-products/vine-vigor-products/

Wikipedia. 2020. Deep Learning. Wikipedia. Referenced 10.5.2020. Available at https://en.wikipedia.org/wiki/Deep_learning

Wikipedia. 2020. Reinforcement learning. Wikipedia. Referenced 11.5.2020. Available at https://en.wikipedia.org/wiki/Reinforcement_learning

Wikipedia. 2020. Supervised learning. Wikipedia. Referenced 10.5.2020. Available at https://en.wikipedia.org/wiki/Supervised_learning