Harnessing coherent illuminator properties to detect subsurface scattering

T. Alder
u7287129

May 29, 2023

Report submitted for **ENGN3712** at the School of Engineering, Australian National University

Australian National University

seeingmachines

Project Area: **Photonics, Image Processing**
Project Supervisor: **Mr. Lachlan Whichello**
Second Examiner: **Distinguished Prof. Richard Hartley**

Abstract

The coherent light output by a laser yields unique interference properties in comparison to incoherent light sources. This project harnesses these unique properties to discern bare skin from other categories present on the face within the constraints defined by a driver monitoring software (DMS) application context. More specifically, the project uses various laser speckle imaging techniques on vertical-cavity surface-emitting laser illuminated images to perform clustering of bare skin. Three distinct methods are investigated: beam profile analysis, laser speckle variation analysis and laser speckle contrast imaging (LSCI). All methods are evaluated and a comparison is made highlighting the respective benefits and drawbacks of each technique in a DMS application context. In the final implementation, a convolutional neural network is trained to use temporal LSCI processed images to effectively classify skin. A comparison of models trained on light-emitting diode (LED) versus VCSEL illuminated datasets is presented and a consistent improvement in classification performance is demonstrated for the VCSEL models. Finally, a comprehensive evaluation is provided into the limitations of this model.

Contents

1 Introduction

A 2014 study by the American Automobile Association concluded some 328,000 crashes annually. 115,400 (35%) of these crashes were attributed to drowsy driving, 109,000 of which resulted in injury and 6,400 crashes which led to fatalities. This accounts for approximately 21% of total annual road fatalities in the United States [1]. In the race to reduce annual road fatalities to zero, some regions are now making driver monitoring systems (DMS) mandatory in all new vehicles. Most recently, in the European Union (EU), the new Vehicle General Safety Regulation came into effect, requiring all new vehicles possess a number of DMS, including attention warnings in case of driver drowsiness or distraction [2]. Furthermore, from 2023 onwards, direct DMS will be required to achieve a five star Australasian New Car Assessment Program (ANCAP) safety rating [3], creating an incentive for automotive original equipment manufacturers (OEMs) outside of the EU to integrate this technology.

In the future, it is expected that such legislation may extend to include occupant monitoring systems (OMS) as well. Projected growth in the DMS and OMS industries is optimistic, with some suggesting the industry will be worth as much as $4203.4 million by 2028 [**?**]. DMS comes in many form factors, however, all direct DMS involve the use of a camera positioned in the automobile cabin for monitoring of the driver (**Figure 1**).

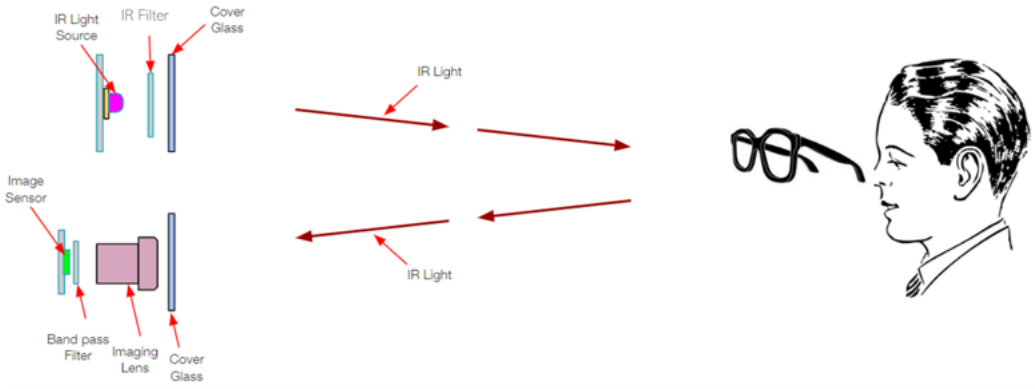

Figure 1: *Exemplar DMS optical pathway.*

In an effort to gain advantage over competitors and achieve a greater percentage of the market share, many DMS/OMS companies are seeking to add additional features to their systems such as monitoring of biometrics signals. One such feature is the ability to detect organic skin. This feature could be useful for spoofing detection or to identify regions-of-interest (ROIs) for monitoring of other biometric signals (e.g. photoplethysmography).

This paper explores the intrinsic interference properties of coherent light and how such properties may be exploited for skin detection (**Figure 2**).

More specifically, it is investigated how vertical-cavity surface-emitting lasers (VCSELs) can be used to perform skin detection. A method of skin classification is formulated using a deep-learning neural network trained on images processed with laser speckle image processing techniques. Results are presented on the efficacy of this neural network in a highly controlled environment and a subsequent effort is made to determine the impact of each control variable on classification performance.

This investigation is performed under the assumptions that an existing DMS system is in place which includes an algorithm that can perform automated segmentation of regions of the face. Constraints of the project are largely derived from the requirement for the final classification method to be practically implementable in an existing DMS system. More specifically, this means that the system should: (1) use near-infrared (NIR) light, (2) be of low cost, (3) be

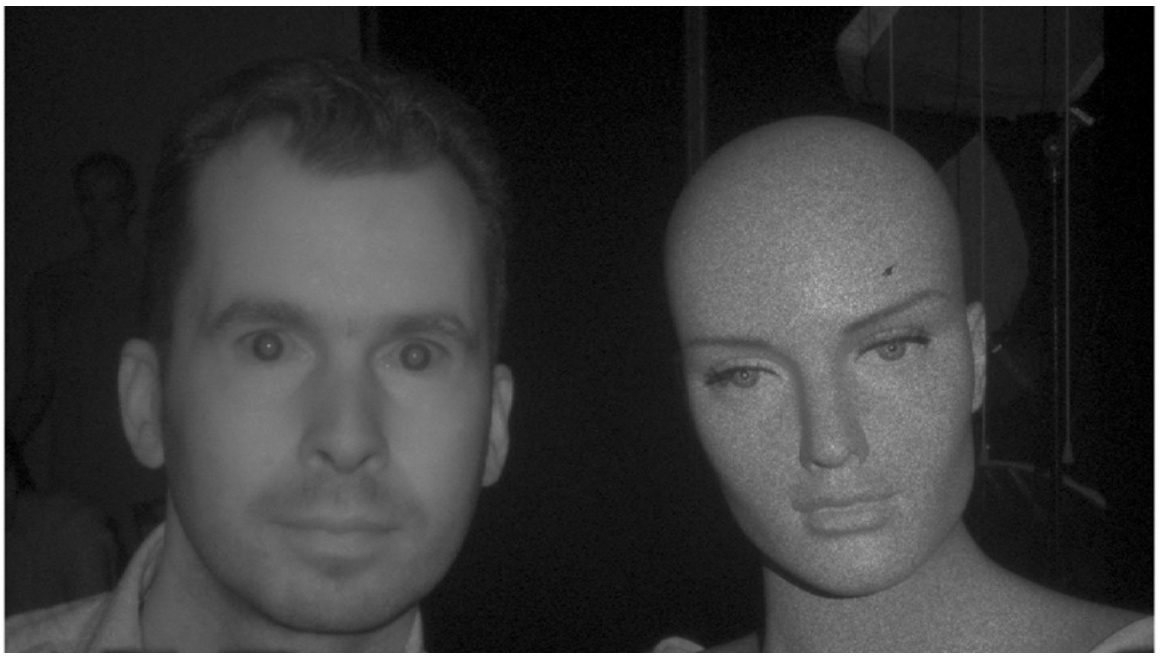

Figure 2: *Coherent light interference as appears on translucent organic skin (left) versus an opaque plastic dummy (right). Notice that the interference operations are much more prevalent on the opaque surface while they are mitigated on the translucent skin.*

of low computational demand, (4) be non-invasive to the user, and (5) require minimal changes to the existing DMS optical pathway. Finally, (6) any changes to the DMS optical pathway should have minimal impact on the image quality of the system.

Thus, the core problem statement of the project is: *can the interference patterns produced by coherent light be used for skin detection within the constraints defined by a DMS application?*

2 Background

Consider two sinusoidal wave forms of the same frequency (**Figure 3**).

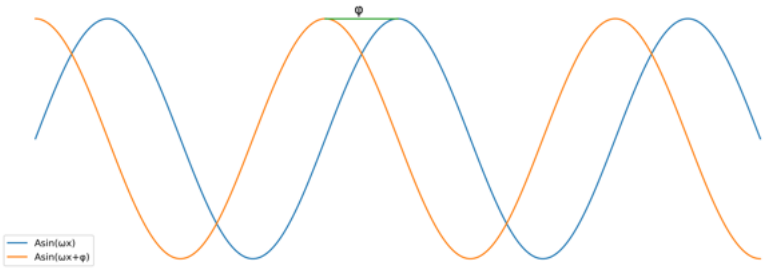

Figure 3: *A plot of two arbitrary sinusoidal waveforms of equal frequency with a phase difference of* φ

In the above case, the relative phase between the two waves would be φ. Two waves of the same frequency are said to be coherent if they have a constant relative phase throughout time. Coherence is significant as it satisfies the necessary preconditions for constructive and destructive interference of two waves. The pre-eminent example of this is Thomas Young's Double Slit experiment (**Figure 4 (a)**).

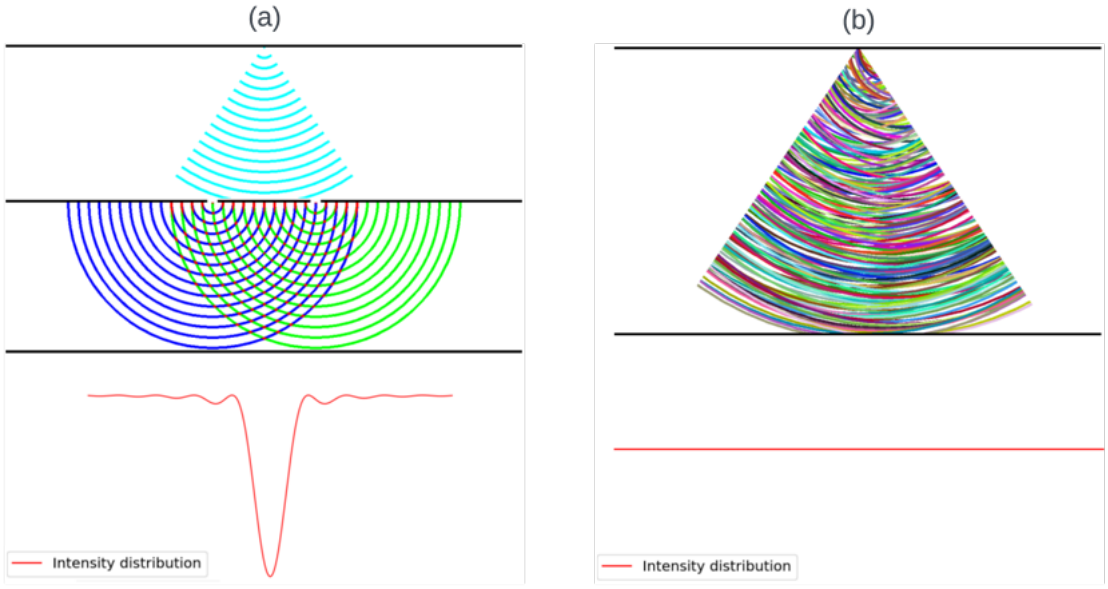

Figure 4: (a) *Example of constructive and destructive interference for a monochromatic, coherent light source as demonstrated by Thomas Young's Double Slit Experiment. The resultant intensity distribution is non-uniform.* (b) *Example of interference operations for an incoherent white light source. The resultant intensity distribution is uniform.*

Comparatively, when a scene is illuminated by incoherent light, the resultant image appears evenly lit due to many operations of interference that average to a uniform irradiance (**Figure 4 (b)**). Thus, as an illuminator, coherent light sources provide exciting opportunities in the realm of signal processing. To appreciate why, one must first have a basic understanding of photonics.

2.1 Photonics

According to quantum electrodynamics, there are three basic actions through which all observed phenomena involving light and matter may be explained: a photon goes from place to place, an electron goes from place to place, and an electron emits or absorbs a photon [5]. Each of these actions takes place in space-time and has a *probability amplitude* which may be calculated.

When a photon crosses a material boundary, there is an electromagnetic interaction with the particles of the medium. This interaction can be characterised by scattering. Scattering of light is primarily explained through action three: the *coupling*, or *junction*, of a photon and electron.

Consider a surface composed entirely of hydrogen. The nucleus of a hydrogen atom consists of a single proton and electron. In the case of scattering, photons are absorbed by electrons orbiting the nucleus and new photons are emitted later in time (**Figure 5**).

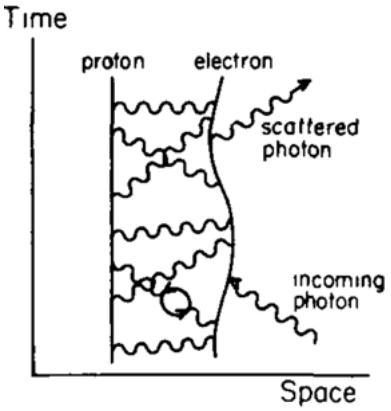

Figure 5: *Example of scattering of a photon by a hydrogen atom. Due to the protons significantly higher mass—in comparison to the electron—it can be considered as stationary in space-time* [5].

There are other ways scattering can occur such as a new photon been emitted before the old photon is absorbed, however, these possibilities can be summed up as a single probability amplitude. The magnitude and direction of this probability amplitude depends on the nucleus and the arrangement of electrons in the atoms, meaning it differs between materials.

Additional characteristics of photon propagation can be defined from the phenomena of scattering, including, but not limited to: transmission (such as through opaque and transparent media), reflection (including partial reflection), absorption and index of refraction.

The manifestation of low-energy interactions between an incident photon and a material depends on the atomic structure of the material. For example, lithium atoms contain three protons and three electrons. The Pauli exclusion principle dictates that it is not possible for more than two electrons with the same spin quantum number to exist at the same point in space-time. As fermions, there are only two quantum states of spin for an electron. This means that the third electron in a lithium atom must be further from the nucleus than the other two atoms, and thus, exchanges few photons. This causes the electron to easily break away from its own nucleus under the influence of photons from other atoms. This phenomena is the basis of conduction. Comparatively, hydrogen and helium atoms do not easily lose their electrons to other atoms. They are *insulators*.

As Feynmann writes in The Strange Theory of Light and Matter, “All atoms are composed of a certain number of protons exchanging photons with the same number of electrons. The patterns in which they gather offer an enormous variety of properties: metals, insulators, gases,

crystals, soft, hard, colored, transparent, etc. A terrific cornucopia of variety and excitement comes from Pauli's exclusion principle and the repetition of three very simple actions" [5].

2.2 Coherent Light Illuminators

As illustrated in **Figure 4 (a)**, coherent light sources produce significant variations in intensity due to operations of constructive and destructive interference. This variation is commonly referred to as *laser speckle* (**Figure 6**).

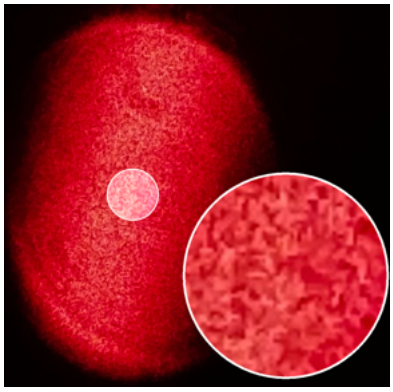

Figure 6: *Laser speckle produced by the constructive and destructive interference of an edge-emitting laser (a coherent light source) illuminating a sheet of paper (a diffuse surface).*

Visually speaking, an image illuminated by a coherent light source is equivalent to a 'noisy' image. For opaque surfaces, the appearance of wave interference is overt, however, such interference is less apparent for translucent surfaces (**Figure 2**). This is due to the nature of light transport in a translucent material; commonly referred to as subsurface scattering.

When coherent light strikes a translucent surface, photons penetrate the surface and are scattered a number of times before passing back out of the material at irregular angles from locations different from the initial point of incidence (**Figure 7**).

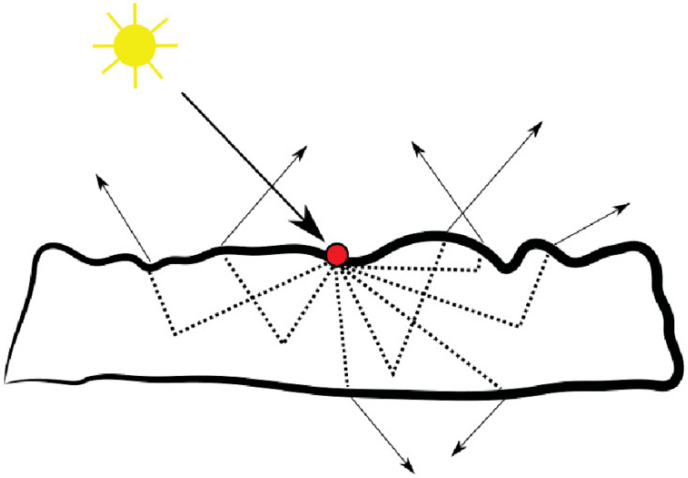

Figure 7: *Subsurface scattering describes the process by which photons penetrate a surface and undergo a number of scattering operations before reflection back out of the materials at irregular angles from locations different from the initial point of incidence* [6].

The effect of subsurface scattering on coherent light sources is equivalent to randomising the phase and polarisation distributions, reducing the presence of wave interference. Accordingly, coherent light sources—such as VCSELs—provide a viable method of discerning between opaque and translucent materials by analysing the presence of subsurface scattering. This paper demonstrates that such differentiation is surprisingly simple using basic image processing techniques, however, fine tuning for more dynamic performance is substantially more difficult.

3 Literature Review

This section seeks to examine existing literature for the purpose of identifying methods of skin detection that could potentially be implemented in a DMS application context. Methods specifically designed for spoofing detection are examined as well as other more trivial methods of skin detection. Research on the topic of skin detection is rather stagnant, however, the field of biometric authentication is rich with research due to significant funding invested into spoofing detection techniques. Accordingly, much of the research examined was derived from papers in the field of biometric authentication and spoofing detection. Some of the research strays from the topic of skin detection for the purpose of ascertaining a wide understanding of relevant photonics phenomenona and image processing techniques. Research is separated into two categories: photonics-based and algorithmic-based detection methods. These categories are slightly misleading as all methods examined involve both photonics and image processing algorithms, however, the distinction serves to discern between methods where detection is heavily reliant on custom optical pathway designs versus methods where a specialised optical pathway is unnecessary.

3.1 Photonics-based Detection Methods

3.1.1 Interferometry

Interferometers are extremely high precision measuring instruments that are commonly used to quantify differences in optical path lengths. The working principle behind interferometers harnesses the interference properties of coherent light. This is accomplished by using a beam splitter to separate a single stream of coherent light into two identical beams. These two beams are then each made to travel a different path length, resulting in a phase difference between them (**Figure 8**). This phase difference produces a static interference pattern much like the 'laser speckle' discussed above. This interference pattern may be used to measure the difference in the optical path length of the two beams.

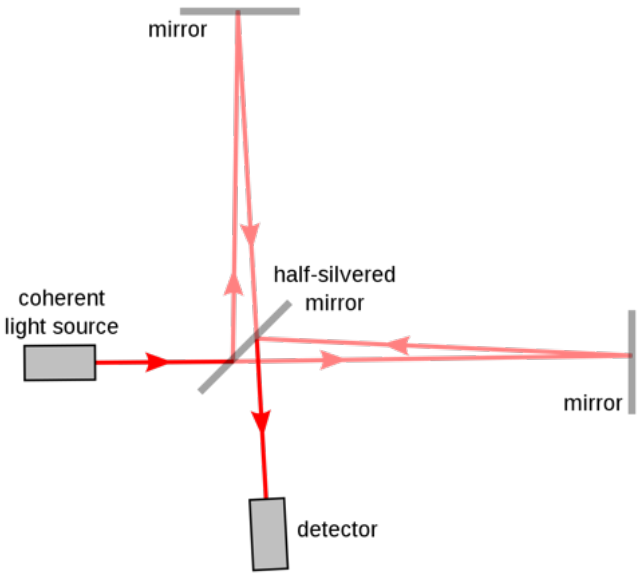

Figure 8: *The experimental setup of a traditional Michelson interferometer* [7].

Furthermore, when a scene is illuminated by the beams, displacement in the shape of the illuminated object results in changes in the intensity distribution of the interference pattern.

Observation of these changes is referred to as *holographic interferometry*. In holographic interferometry, an image is taken defining the initial state of the scene and any images taken subsequently can be superimposed over the initial state to examine surface displacement. If the scene in question is a person, holographic interferometry can be used to observe dynamic changes stemming from blood flow and perform vascular mapping [8]. Furthermore, holographic interferometry can be extended to a technique referred to as *laser doppler flowmetry*, whereby the flow rate of the imaged veins may also be determined [9]. The primary limitation of this technique is it requires the subject to remain extremely still between images. In practice, this is difficult to accomplish outside of laboratory conditions. Additionally, the experimental setup required to perform holographic interferometry is costly and complicated. Thus, while holographic interferometry could present an effective method of skin detection, it is not considered to meet the constraints enforced by a DMS use-case.

In a similar experimental process known as Fourier transform spectroscopy, an interferometer can be used to perform emission spectrum analysis [10]. In this case, the same experimental setup described above is maintained, except the difference in optical path length of the two beams is now controllable. This is accomplished through the use of a translating mirror (**Figure 9**).

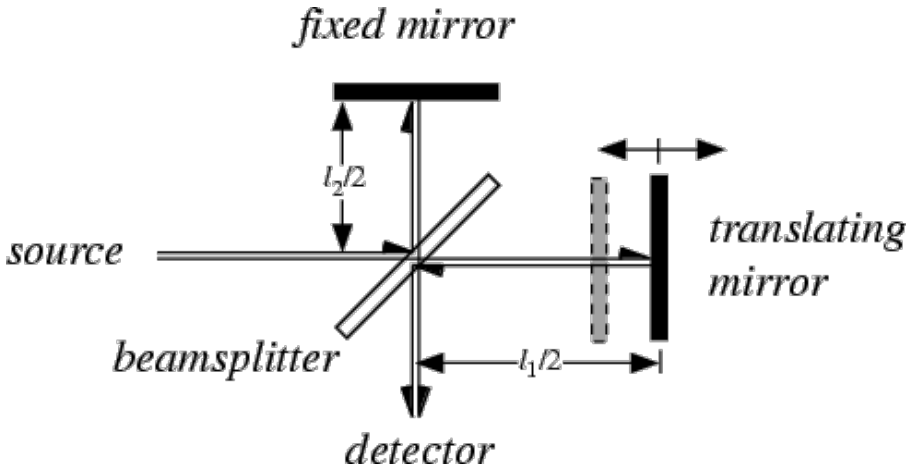

Figure 9: *Experimental setup for Fourier transform spectroscopy. A typical interferometry setup is used except the phase difference between the two beams is now directly controllable through a translating mirror* [11].

Depending on the path lengths travelled by either beam, the subsequent interference operations result in a different range of wavelengths passing to the detector. A number of measurements are taken using difference different path length configurations. The measurements taken from each of these experimental configurations can be combined to form an interferogram (a pattern formed by the interference operations). A Fourier transform is applied to this interferogram to generate the emission spectrum of the light received at the detector. By placing a sample infront of the source or detector, the absorption spectrum of the sample can be observed. Some examples of this spectroscopy technique include [10] and [12].

As elaborated on **Section 3.1.3**, such a spectrum can be used to analyse the atomic and molecular structure of the sample. Accordingly, Fourier transform spectroscopy presents a robust, viable method of skin detection. However, Fourier transform spectroscopy suffers from the same limitations as holographic interferometry when implemented in a DMS context. Thus, it is not considered to be a viable method of skin detection for a DMS application scenario.

3.1.2 Beam profile analysis

A 2020 white paper published by German-based company trinamiX GmbH detailed the use of a coherent light dot projector illuminator paired with a high-resolution complementary metal-oxide semiconductor (CMOS) image sensor for the purpose of material detection [13]. trinamiX accomplish this by analysing the intensity (beam profile) of each imaged projected dot. Accordingly, the quality of the signal sampling is influenced by the physical CMOS resolution, the size of the projected dots and the camera's objective lens. The imaged projected dot is analysed using a number of convolution kernels that extract both distance (the beam profile intensity varies with distance) and material classes. For further refinement of material classification, trinamiX state that their beam profile analysis is combined with deep learning strategies. Through such processes, trinamiX's method is capable of discerning between a wide variety of materials including: skin, clothing fabric, hard plastic and seat belt webbing [14].

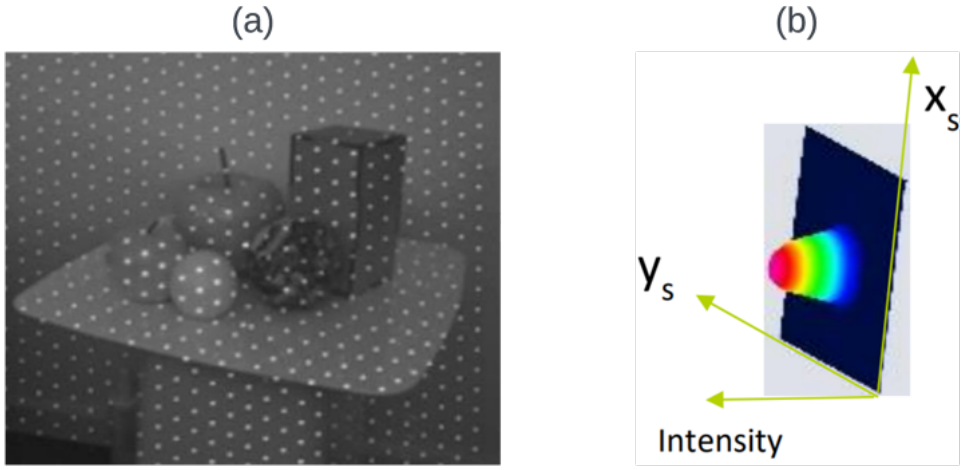

Figure 10: (a) *An example of a scene illuminated by a coherent light dot pattern projector. Notice that the intensity of each projected dot is different depending on the material of incidence* [13]. (b) *An example of the intensity profile of a projected dot. The intensity distribution can be modelled as a Gaussian where the peak is dependent on the material of incidence and the depth of projection distance* [13].

trinamiX have a working, commercially available product associated with this patent. trinamiX's product uses 2000 dots projected across a scene with time regimes of 100μs-2ms and a further 48,000 dots are created through "smart interpolation". With these dots, they are able to perform accurate material classification that is robust in ambient light for distances up to 1m [13].

Through observation, it can be seen that the efficacy of trinamiX's implementation of this method of material classification wanes at edges separating materials with similar optical properties (e.g. seatbelt webbing and clothing fabric). Unsuprisingly, classification performance is good for materials with a large contrast in optical properties (e.g. skin versus hard, opaque plastic). A patent filed by trinamiX on this technique indicates they have a binary classification system (i.e. skin and not-skin). However, promotional material released by trinamiX indicates they are capable of a wider variety of classification in specific environments such as within a car cabin (**Figure 11**).

This technique of skin detection is investigated in the methodology of this report (**Section 5.1**). A simple thresholding operation of the variance of each beam profile was used to cluster

Figure 11: *An example of trinamiX GmbH beam profile analysis material classification technique deployed in a DMS application scenario* [14].

projected dots, producing impressive results. Unfortunately, however, further investigation of this technique was not completed due to hardware limitations.

3.1.3 Spectra analysis

As discussed in **Section 2.1**, when light is directed at a material, a proportion of the incident light is absorbed, while the remainder is transmitted and/or reflected. Assuming an idealised full-spectrum illumination source is used, absorption is concentrated in wavelengths dependent on the atomic composition of the imaged material (**Figure 12**). This phenomenon is the basis of spectral line analysis, which is used by astrophysicists to identify the atomic and molecular components of distance stars and planets.

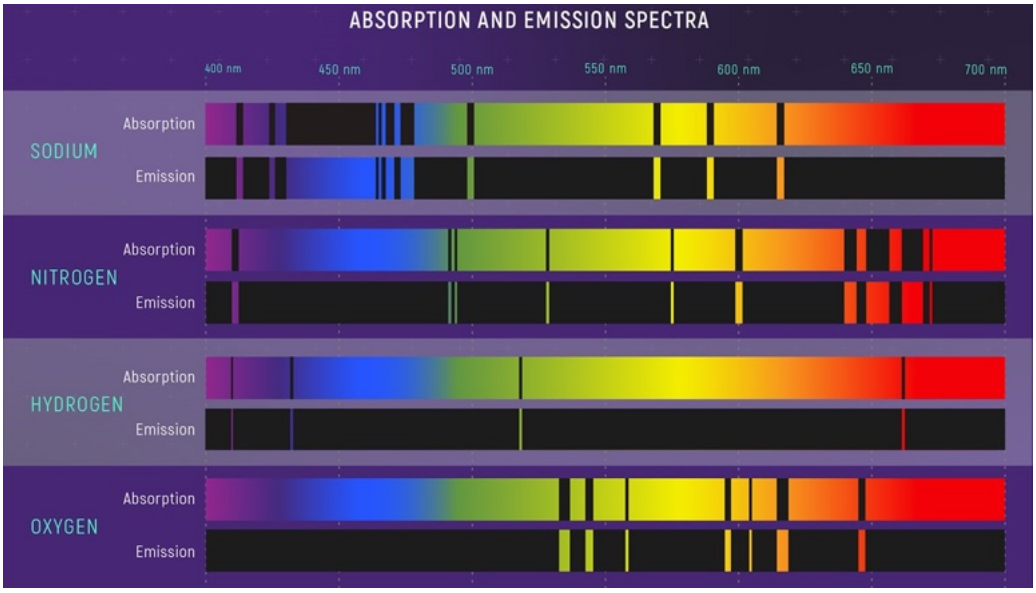

Figure 12: *The emission and absorption spectrum's of sodium (Na), nitrogen (N), hydrogen (H) and oxygen (O) for the visible light spectrum* [15].

Such analysis can be performed using an optical absorption spectrometer. Accordingly, an optical absorption spectrometer could be used to classify skin, as well as any other materials present in a scene. Although this would result in an extremely versatile and robust method

of skin detection, these benefits come at a high economic cost (several thousand dollars for a single spectrometer). Accordingly, it is economically unfeasible to use this technique in a DMS application.

However, given the projects scoping constrains the problem statement to just skin detection, the detection method of spectra analysis is not entirely invalidated. For example, a 2008 patent filed by Honeywell International Inc [16] uses just two narrow-bandwidth NIR sources for skin classification. The first NIR source has a bandwidth of 800nm-1400nm and the second IR source has a bandwidth of 1400nm-2200nm. Due to the characteristics of photon-skin interaction (**Section 4.1**), these bandwidths mean that the first NIR source is reflected by skin while the second NIR source is absorbed. The reflected spectrum of the imaged surface is then analysed and the characteristics of the absorption spectrum are used to determine whether the imaged surface was skin.

In essence, this is a primitive spectrometer. However, it is important to note that many materials other than skin reflect light in the 800nm-1400nm range and absorb light in the 800nm-1400nm range. Accordingly, while this method may present a viable method of skin detection in highly contrasting environments, it is unlikely it would prove effective when presented with a scene of materials with optical properties similar to that of skin (e.g spoofing attempts).

3.1.4 Laser speckle image processing

As discussed in **Section 2.1**, when a coherent light source interfaces with a diffuse surface, an interference pattern—commonly referred to as *laser speckle*—is produced. The produced pattern is static, however, any moving particles inside the imaged ROI cause fluctuating, dynamic speckle patterns related to the flow dynamics present within the structure. These dynamics can be revealed through the application of laser speckle imaging (LSI). More specifically, temporal or spatial processing of LSI images can be used to increase the contrast between static and dynamic regions of speckle. LSI processing is commonly utilised by medical professionals as a method of non-invasive vascular imaging [17]. In this application context, the primary limitation of LSI processing is the requirement of exposed vascular tissue (i.e. it cannot be used to image subcutaneous blood flow).

A 2020 paper by Dolan *et al.* investigates the viability of three distinct LSI processing techniques: Laser Speckle Contrast Imaging (LSCI), Generalised Differences and the Fuji method [18]. The implementation of these techniques is detailed in **Figure 13**.

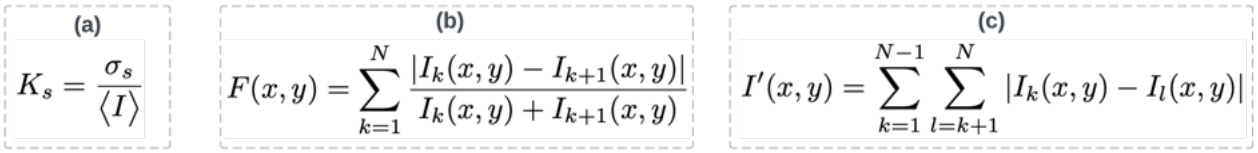

Figure 13: *LSI image processing techniques employed by Dolan et al.* [18]. (a) *Temporal or spatial LSCI computation,* (b) *temporal Fuji computation, and* (c) *temporal Generalised Differences computation.*

In the case of LSCI computation (**Figure 13 (a)**) the ratio of standard deviation and median intensity is computed. This computation may be done spatially (using a neighbourhood of pixels), or temporally (using the same pixel index across a collection of images). The resultant value—the speckle contrast value—is between 0 and 1 with 1 indicating totally static conditions with no motion and 0 indicating a fully 'decorrelated' speckle pattern caused by the fast motion of scatterers. In the case of temporal Fuji computation (**Figure 13 (b)**), a weighted sum of the absolute difference between frames of associated pixels is computed. Dolan *et al.* note that

this method is susceptible to regions of low-illumination, which could be treated as areas of high activity. Finally, in the case of Generalised Differences computation (**Figure 13 (c)**), variation between non-consecutive frames is considered. The result is similar to that of the Fuji method, except slow moving dynamics that are not distinguishable between consecutive frames are also highlighted.

The authors apply each of these techniques to an image of a Manhattan Euonymus leaf illuminated by a NIR laser, captured using a monochromatic CMOS sensor. 1500 images are captured across 50 seconds with a resolution of 480x480 pixels. The authors findings are presented in **Figure 14**.

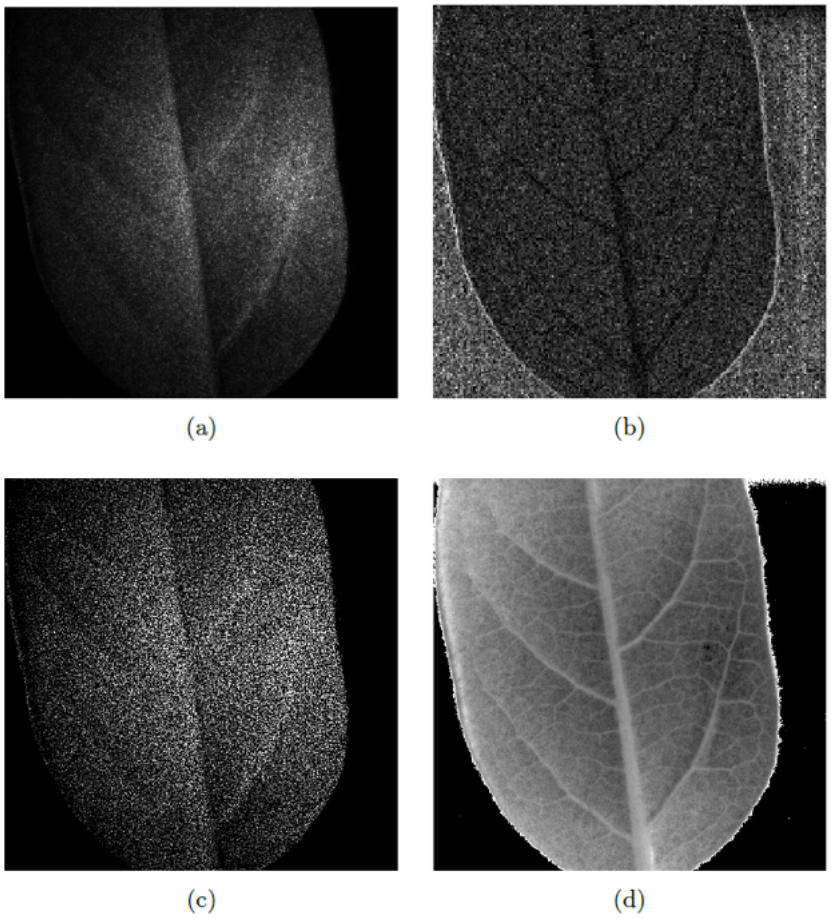

Figure 14: (a) *Raw speckle image of Manhattan Euonymus leaf and images obtained with* (b) *LSCI,* (c) *Generalised Differences and,* (d) *the Fuji method* [18].

The authors finding suggest that the Fuji method is, by far, the most effective LSI image processing technique for increasing contrast of the vascular structure of the leaf.

With regard to skin detection, this use-case of this technique is very similar to that which is described for holographic inteferometry (**Section 3.1.1**). More specifically, LSI image processing could be used to perform vascular mapping of an ROI and a classification algorithm could then be used to ascertain whether a vascular map was in fact produced (indicating the presence of organic tissue - i.e. skin). The primary difference between these two methods is the complicated experimental configuration required to perform interferometry. Comparatively, LSI image processing uses a vastly simplified and more economical experimental setup. The trade-off for this is an inability to determine volumetric flow rate. However, volumetric blood flow rate is outside of the scoping of this project. Thus, from the context of a DMS skin detection application, LSI image processing addresses the limitations of holographic interferometry. The application of LSI LSCI processing is investigated in the methodology of this report, although

its usage is distinct from that presented by Dolan *et al.* in that it is not used for vascular mapping. Instead, LSCI is combined with a number of other image processing techniques to increase the textural contrast within consecutive source images. For more information on this application, see **Section 5.3**.

Another method of LSI image processing is laser speckle variation analysis. As discussed in **Section 2.1**, when a coherent light source interfaces with a diffuse surface, an interference pattern—commonly referred to as *laser speckle*—is produced. This effect is mitigated when the incident surface is translucent due to the presence of subsurface scattering. In laser speckle variation analysis, a raw LSI image is taken and variance statistics are computed for neighbourhoods within the image. Each pixel may then be clustered according to the variance statistics of it's neighbourhood. Further insight into this technique can be derived from a 2018 patent filed by California-based company, Lumileds [19]. Lumileds, uses this technique for the creation of a biometric authentification system primarily designed for application in mobile phone devices. The patent defines two distinct sources of speckle: 'illumination speckle' and 'imaging speckle'. Imaging speckle is the aforementioned *laser speckle* inherent in coherent light sources. Comparatively, illumination speckle is a static diffraction pattern produced by the passing of a laser beam through a surface and/or volume diffusor. Lumileds use variation in illumination speckle to cluster different materials present within a scene. Lumileds claim their method is capable of detecting variation in speckle contrast of at least 1%. Lumileds also claim their method is implementable with NIR coherent illuminators through the use of a bandpass filter or with modulated RGB-NIR illuminators through the use of image subtraction. For optimal performance, Lumileds state the detecting camera should have an aperture of 1.5-2.5mm (a larger aperture results in a less-pronounced speckle pattern) and the illuminator should be no more than 15mm from the lens for even illumination of the subject. Despite these claims, it is important to note that no specific algorithmic image analysis techniques are defined in the Lumileds patent nor does any associated commercially available product exist on Lumileds website. This suggests they have patented the 'idea' of biometric analysis using laser speckle rather than a specific method for accomplishing this.

In regards to skin detection, laser speckle variation analysis could be used to cluster regions of skin within a scene. Presumably, this technique could be used to produce a robust algorithm capable of skin classification using the observed variation of laser speckle as an identifier. Laser speckle variation analysis is attempted in the methodology of this report, although initial findings showed little promise for use as a method of skin detection. It is important to note, however, the implementation of this technique specific conducted in this report is non-ideal, due to the constraints presented by a DMS application context. Theoretically, with a less constrained project scope, this technique could be used as a viable form of skin detection. For more information on the implementation of laser speckle variation analysis attempted in this report, see **Section 5.2**

3.2 Algorithmic-based detection methods

3.2.1 Bidirectional subsurface scattering reflectance distribution function

In 2001, Jensen *et al.* released their seminal paper 'A Practical Model for Subsurface Light Transport' [20]. Prior to the papers release, the primary technique for simulating reflectance in computer graphics was the the bidirectional reflectance distribution function (BRDF) model. *Jensen et al.'s* paper overcame fundamental limitations within this model, allowing for the capture of color bleeding within materials and diffusion of light across shadow boundaries. In the same paper, Jensen *et al.* also introduce a rapid image-based measuring technique for determining the optical properties of translucent materials such that they may be simulated. This technique could be used to extract the optical properties of skin from a dataset of images.

Theoretically, a classification model could be trained to perform this task and use the extracted properties for skin detection.

The variables used by Jensen *et al.* to model the optical properties of translucent materials are a reduced scattering coefficient (σ'_s), an absorption coefficient (σ_a), and a diffuse reflection coefficient. For each colour channel, a different value is used for these coefficents. Thus, in a three-dimensional colour space, one material will have nine associated BSSRDF coefficients. Some examples of these parameters recorded for a range of materials by can be seen in **Figure 15**.

Material	σ'_s [mm^{-1}]			σ_a [mm^{-1}]			Diffuse Reflectance			η
	R	G	B	R	G	B	R	G	B	
Apple	2.29	2.39	1.97	0.0030	0.0034	0.046	0.85	0.84	0.53	1.3
Chicken1	0.15	0.21	0.38	0.015	0.077	0.19	0.31	0.15	0.10	1.3
Chicken2	0.19	0.25	0.32	0.018	0.088	0.20	0.32	0.16	0.10	1.3
Cream	7.38	5.47	3.15	0.0002	0.0028	0.0163	0.98	0.90	0.73	1.3
Ketchup	0.18	0.07	0.03	0.061	0.97	1.45	0.16	0.01	0.00	1.3
Marble	2.19	2.62	3.00	0.0021	0.0041	0.0071	0.83	0.79	0.75	1.5
Potato	0.68	0.70	0.55	0.0024	0.0090	0.12	0.77	0.62	0.21	1.3
Skimmilk	0.70	1.22	1.90	0.0014	0.0025	0.0142	0.81	0.81	0.69	1.3
Skin1	0.74	0.88	1.01	0.032	0.17	0.48	0.44	0.22	0.13	1.3
Skin2	1.09	1.59	1.79	0.013	0.070	0.145	0.63	0.44	0.34	1.3
Spectralon	11.6	20.4	14.9	0.00	0.00	0.00	1.00	1.00	1.00	1.3
Wholemilk	2.55	3.21	3.77	0.0011	0.0024	0.014	0.91	0.88	0.76	1.3

Figure 15: *Some BSSRDF modelling coefficients generated by Jensen et al. [20]. Note that η is the relative index of refraction for each material.*

In regards to skin detection, the findings of Jensen *et al.* presented in **Figure 15** demonstrate how the measurement of BSSRDF coefficients could be used for material classification. It is important to note, however, that different coefficients are given for different colour channels. In a three-dimensional colour space, this provides a high degree of variance between materials with similar optical properties. In a monochromatic colour space this variance would be significantly decreased and, thus, it is likely this method would struggle to distinguish materials with similar optical properties. Additionally, computation of the BSSRDF coefficients performed by Jensen *et al.* was highly dependent on the positioning of the illumination source relative to the imaging plane. The authors note that Lambertian reflectance in conjunction with a Fresnel scaling term can be used to account for changing in positioning. In a controlled environment, this achievable, however, in a DMS application context, ambient sunlight would render this calculation virtually impossible. Accordingly, this method is not considered to be a viable method of skin detection.

3.2.2 Pixel-based detection

Pixel-based skin detection refers to the application of image processing techniques for the identification of skin pixels in an image. This is accomplished by translating the raw image to a suitable colour space and creating (or training) a classification algorithm that labels pixels as skin or not-skin. Research on this topic is extensive due to its application for fast filtering of adult images on the world wide web. Many authors focus on the impact of colour space on the performance of an algorithm, however, in 2001, Albiol *et al.* mathematically proved there exists an optimal detection algorithm for each colour space implying performance is independent of colour space [21].

Numerous methods of classification have been investigated under this topic. Some of these include: parametric models (e.g. Gaussian mixture models), non-parametric (histogram-based) models [22], artificial neural networks, spatial analysis methods and adaptive methods [23].

The primary benefit of pixel-based detection methods is their speed. The simplistic nature of the image processing techniques employed by pixel-based detection methods means their computational efficiency is largely unparalleled. However, this efficiency is not without trade offs; pixel-based detection is generally less accurate than other detection methods and far more susceptible to spoofing.

While pixel-based skin detection does not meet the criteria of this project (mainly in regards to robustness), it could be used as an efficient method of masking potential ROIs (i.e. regions of skin) from a source image. Other more computationally intensive techniques could then be applied to these ROIs to determine whether they are in fact skin.

3.2.3 Machine learning applications

At the core of this reports problem statement is a classification problem: *how can one use the presence of coherence light interference to classify regions of skin?* Over the past decade, machine learning has seen exponential progress in its application to classification problems. This is chiefly due to advancements in supervised learning; more specifically, transfer learning. The machine learning community has now progressed to a point where a user may take an existing pretrained model, fine-tune it on a dataset of images specific to their use-case, and achieve state-of-the-art classification results. A good example of this is detection of skin diseases such as melanoma [24] and other more common dermatological diseases [25].

As a good rule of thumb, a machine learning model can generally perform well on any classification problem that is distinguishable to the human eye. This philosophy underpins the core of the methodology of this project, where image processing techniques discussed in this literature review are applied to images illuminated by coherent light until contrast between different classes is increased to a degree where it may be distinguished using the human eye. A deep learning classification learning network is then fine-tuned on these images. The results of this classification algorithm are presented in **Section 6.5** and **Section 7**.

4 Optical Properties

4.1 Skin

To develop a robust classification algorithm capable of discerning skin in a variety of environments, it is important to have an informed understanding of the unique optical properties of skin. The intention of this section is to develop this understanding, such that a classification algorithm can be developed that harnesses the identified optical properties.

Interpreting the interaction of photons within skin is a complex and arduous process stemming from its composition of many layers of distinct organic materials. In order to accurately model the optical pathway of light through skin, a combination of scattering models are required. For example, the case of visible light is presented in **Figure 16**.

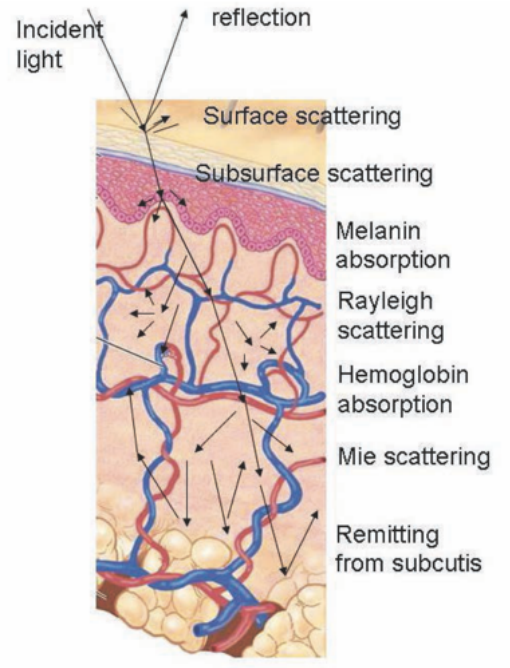

Figure 16: *The complex optical pathway of visible light through skin can be modelled using a combination of scattering models* [26].

Furthermore, the characteristics of this interaction varies depending on the wavelength of the incident light. This is why veins appear blue: because photons of wavelength 450-495nm (the blue portion of the visible light spectrum) do not penetrate as deeply as lower energy photons such as the red (620-750nm). Broadly speaking, mammalian skin is comprised of two layers: the epidermis and the dermis.

4.1.1 Epidermis

The epidermis refers to the outermost layers of skin. In order of outermost to innermost, these include: the stratum corneum, the stratum lucidium, the stratum granulosum, the stratum spinosum, and the stratum germinatum. The epidermis is continually regenerated, with new cells formed in the stratum germinatum (innermost layer) and slowly migrating to the outermost layer as they undergo a maturing process referred to as keratinization.

In the epidermis, the primary photon interaction is absorption. This is due to the presence of the amino acid melanin [27]. The various hues and degrees of pigmentation found in the skin of human beings are directly related to the number, size, and distribution of melanosomes (the

organelle responsible for synthesising melanin) within the epidermis. Jacques [28] estimates that, in light-skinned adults, 1.6-6.3% of the epidermis of comprised of melanosomes, while, in darkly pigmented adults, 18-43% of the epidermis is comprised of melanosomes. Accordingly, it is expected that laser speckle will be diminished within darker skinned individuals due to the increased absorption of photons stemming from a greater degree of melanin.

The fibrous protein keratin is also responsible for some light scattering in the epidermis, however, its impact is negligible in comparison to melanin.

4.1.2 Dermis

Photon interaction within the dermis is largely characterised by scattering. More specifically, the presence of the fibrous protein collagen is the primary source of light scattering in the dermis. These scattering interactions can be modelled as Mie and Rayleigh scattering [28]. The hemoglobin protein within red blood cells also contributes to scattering within the dermis.

The scattering and absorption operations within the epidermis and dermis act to randomise the phase and polarisation distributions of coherent light incident on skin. This results in the diminishing of laser speckle interference pattern. It is this diminishing of laser speckle interference that the methodology of this project will attempt to use for skin classification.

4.1.3 Near-infrared light

The analysis presented above is largely specific to the interaction of visible light with skin. For these shorter wavelengths, photon-surface interaction is largely characterised by absorption by water and melanin. For longer wavelengths such as NIR light, deeper penetration under the skin is observed and the probability of absorption diminishes (**Figure 17**). A benefit associated with this deeper penetration is that it becomes significantly harder to fake (i.e. spoof) a material that will exhibit the same optical characteristics. This also means that the resulting reflectance spectral features are mainly affected by scattering interactions [26]. This can be seen in **Figure 17**, where the absorption coefficient decreases hundredfold for wavelengths 800nm-1000nm, while scattering only decreases fivefold. Note that for wavelengths > 1100nm, the probability of absorption passes a local minima and begins increasing. Comparatively, the scattering coefficient plateaus for these wavelengths.

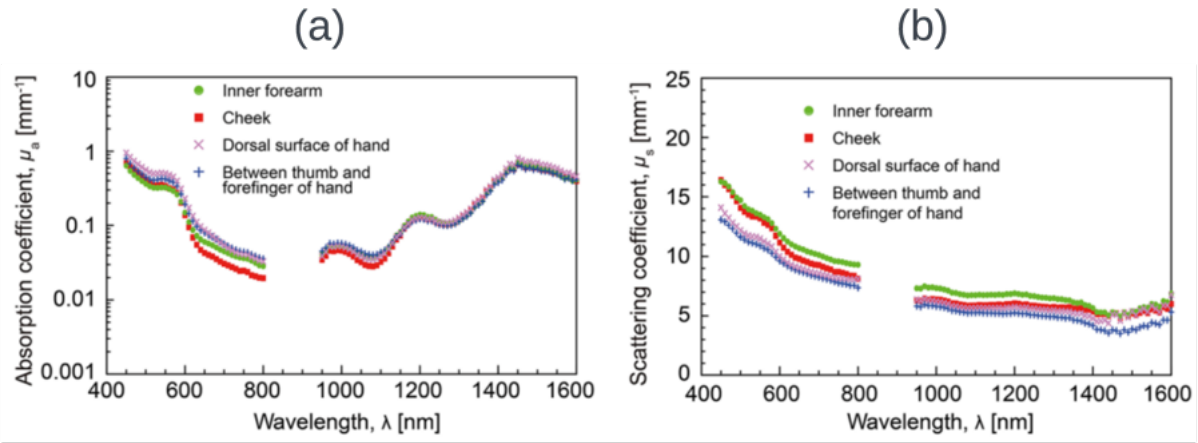

Figure 17: (a) *Absorption coefficient versus wavelength for various regions of skin.* (b) *Scattering coefficient versus wavelength for various regions of skin* [29].

4.2 VCSELs

As illustrated in **Figure 1**, a traditional DMS system consists of one—or several—IR light sources. Traditionally, a NIR illumination profile (with peak emission at 850nm or 940nm) is used due to the drop in the emission spectrum of sunlight at these wavelengths. Conventionally, light emitting diodes (LEDs) are the chosen IR light source for these systems due to their low-cost, high-efficiency, and wide availability.

In recent years, a new form of illuminator is seeing increasing popularity: VCSELs. The internal structure of a VCSEL is pictured in **Figure 18**. Traditional lasers are edge-emitting, which makes reliable mass production wasteful and expensive. In comparison, the techniques used in the manufacturing of surface-emitting lasers have surpassed these limitations and are now in widespread use amongst devices including smartphones, time-of-flight sensors, spectral sensors, and various optical communications technologies. Though VCSELs are currently limited by their cost of production, many academics in the field of optics see them as a potential future alternative to LEDs. This is due to the many exciting applications that are only possible using coherent light sources. Some existing proven examples include: speckle interferometry, 3D depth sensing, optical communication, etc.

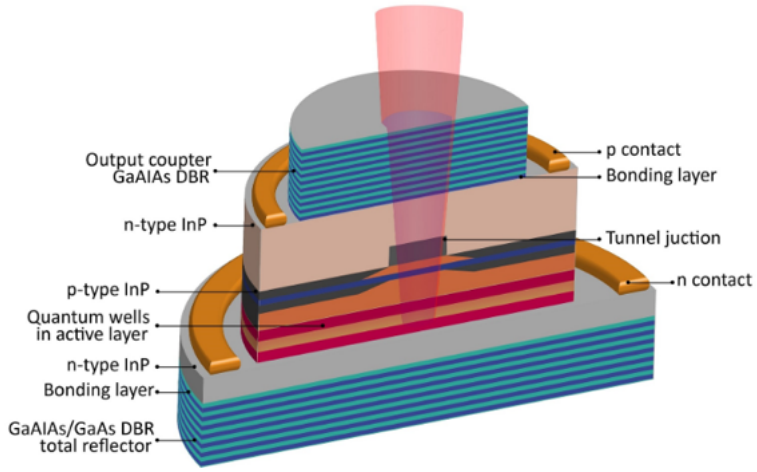

Figure 18: *The internal structure of a VCSEL* [**?**].

The intensity profile of an individual laser emitter is far too small in diameter to evenly illuminate an entire scene. To overcome this limitation, it is common for hundreds of laser emitters to be combined to form an array that is passed through a diffusor and diverging lens in a configuration referred to as a *flood illuminator*. The same interference principle observed in Youngs Double Slit experiment (**Figure 4**) can be extrapolated to a flood illuminator, with the key difference being that interference is observed between hundreds of coherent light emitters (**Figure 19**), rather than just two. At present, to ensure a stable phase relationship (i.e. coherence) is maintained for imaging application scenarios, the number of emitters in a flood illuminator is restricted by manufacturing limitations to order of magnitude 10^2. This limit is driven by Excluding work conducted on beam profile analysis (**Section 5.1**), all VCSEL-based work completed in the methodology of this report used a flood illuminator as the illumination source.

One of the benefits associated with VCSELs is that, through careful selection of additional optical components, virtually any desired intensity profile may be created. For example, it is common for an aspherical lens to be used in conjunction with a beam splitter to create a discontinuous intensity distribution. This is commonly referred to as a *dot project* (**Figure 20**).

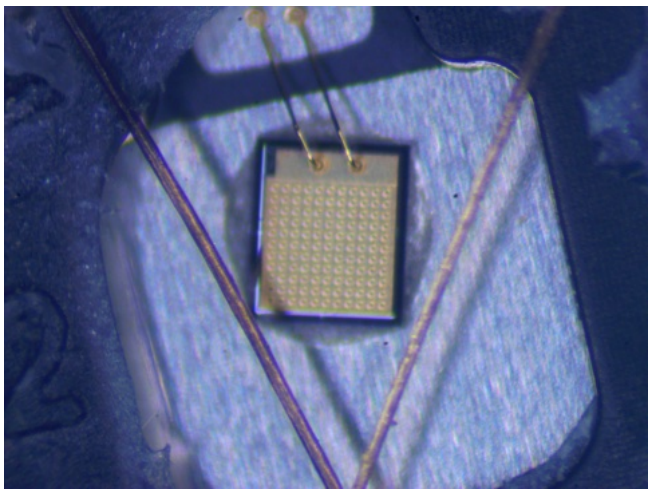

Figure 19: *A VCSEL emitter array with strands of hair for scale. The output of the emitter array is passed through a diffusor and a diverging lens to produce an intensity profile similar to that of an LED.*

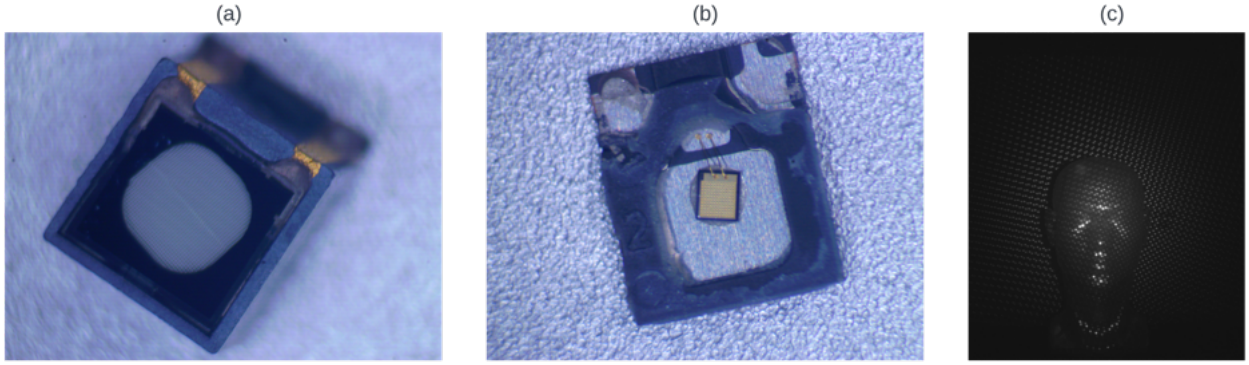

Figure 20: (a) *A dot-pattern projector.* (b) *The same dot-pattern projector with the diffusive cover removed.* (c) *An example of a scene illuminated by an NIR dot pattern projector. Notice that the light is focused in discontinuous regions throughout the scene.*

Dot pattern projectors are the basis behind Apples Face ID technology. As presented in **Section 3.1.2**, dot pattern projectors can also be used for material classification. This technique is investigated within the methodology of this report (**Section 5.1**).

VCSELs have greater efficiency, higher optical power output, narrower bandwidth, less latency, greater temperature stability and greater customisability of illumination profile than LEDs. These benefits come at the detriment of increased cost and worse image quality (due to laser speckle). Accordingly, significant research is focused around how laser speckle in the output of VCSELs may be mitigated.

The first, and most obvious solution to this is broadening linewidth. This may be accomplished by heating up the VCSEL cavities or redesign of the semiconductor. Another approach to broadening linewidth is through a carefully designed diffusive element. Finally, the VCSEL emitter array may be driven in a multi-transverse mode emission regime to reduce the amount of mode overlap (or even into an incoherent regime where transverse modes break down and individual emitters support multiple beamlets) [31]. This represents the fundamental limit on the possibility of eliminating speckle with VCSELs, but requires sub microsecond pulsing for effective implementation. Additionally, existing work in this area only deals with single emitter, broad area VCSELs. It is unclear whether this effect is possible with VCSEL arrays consisting

of hundreds of individual emitters.

5 Material Detection

Several viable methods of skin detection were identified in the reports literature review. Attempts were made to implement three of these methods; laser speckle variation analysis, beam profile analysis, and laser speckle contrast imaging; before the final classification technique was developed. In this section, precursory attempts are made to implement these techniques within the constraints defined by a DMS application context, establishing the logic upon which the final classification algorithm was constructed.

5.1 Beam Profile Analysis

As established in the literature review, it is possible to use laser beam profile analysis to perform classification of both inorganic and organic materials. Hardware specifications are of significant importance for this technique due to an inherent sensitivity to image resolution, image quantisation, beam profile intensity, distance to subject and ambient light intensity. Accordingly, the first step in implementing a beam profile analysis material classifier is developing an appropriate set of hardware specifications. Fortunately, such specifications can be derived from a 2020 whitepaper published by trinamiX GmbH [13].

Table 1: *Beam profile analysis hardware requirements derived from trinamiX GmbH whitepaper* [13].

Hardware Requirements	Point Resolution	Point Density	Illumination Characteristics
Global shutter camera + dot projector	<∼24x24 Px	2k projected + 48k interpolated	850nm or 940nm collimated light

At the time of investigation, a system meeting these hardware specifications was not readily accessible. However, a camera system providing ∼12x12 pixels resolution per dot was available. As a result, work commenced on implementing this technique using the readily accessible sub-optimal hardware.

Firstly, the centroid of each projected dot needed to be isolated such that analysis could be performed for each beam profile. To accomplish this, the source image was upsampled using a Gaussian image pyramid. This allowed for a greater number of projected dots to be isolated. The upsampled image was 'denoised' using a Gaussian filtering operation followed by a median filtering operation. The resultant 'denoised' image was then subtracted from the source upsampled image. The output image consisted of only the dots in the source image (**Figure 21**).

The 'dot image' is then binarised and convolved with another median filtering operation to remove any lingering background signal. Next, OpenCV's *findContours* function is used to isolate each dot and the *moments* function is used to find the centroid of each isolated dot. Finally, a linear transformation operation is used to convert these centroids to the coordinate scale of the original images resolution. In well-lit distances up to a metre away, this technique is effective in isolating 95% of all projected dots, however, this performance can vary substantially depending on illumination conditions and distance to target.

Using the computed centroids, ROIs can be masked around projected dots from the source image. These ROI's can be used to perform beam profile analysis of each projected dot. In the most primitive case, the standard deviation of each beam profile can be used to cluster materials present in the scene. For example, in the case of an arm positioned infront of medium-density

Figure 21: *Masked dots image. Virtually all image content is zeroed excluding the beam profile of each dot.*

fibreboard (MDF), clustering projected dots using a simple binary thresholding operation yields impressive results (**Figure 22**).

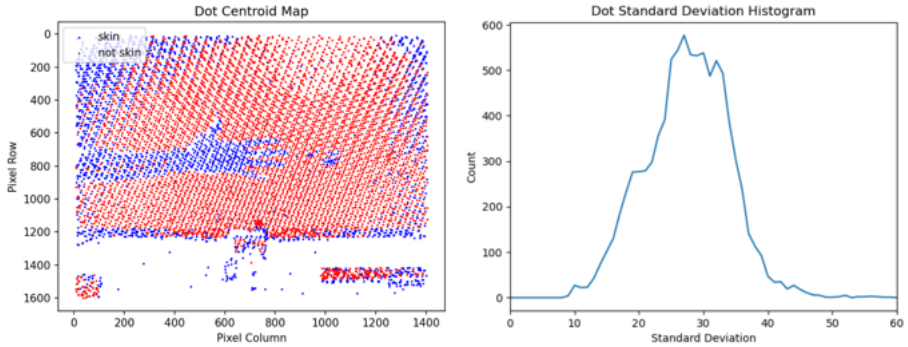

Figure 22: *Beam profile clustering using a binary threshold of the standard deviation of each beam profile.*

This performance can be improved with a multiplicative corrective matrix derived from the cosine fourth power law to rectify diminishing irradiance of projected dots at the edge of the image scene (**Figure 23**)).

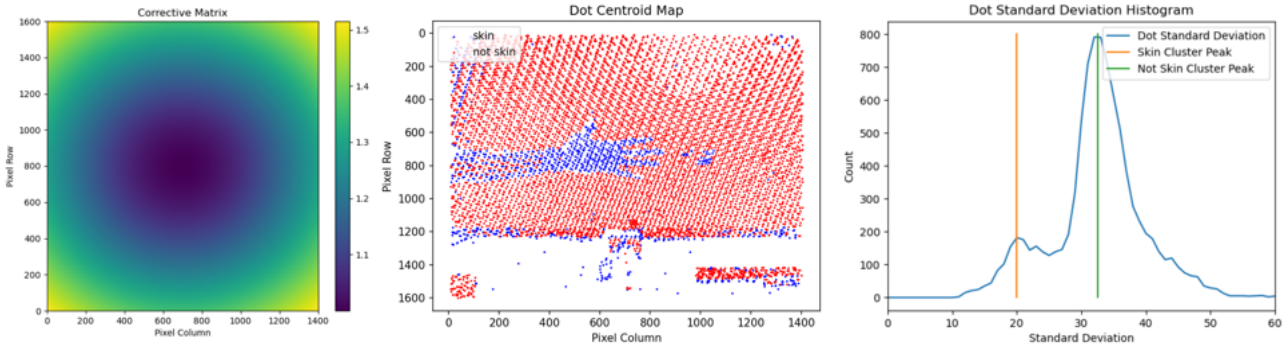

Figure 23: *Beam profile clustering using a binary threshold of the standard deviation of each beam profile with vignetting of the raw input image minimised using a corrective matrix.*

As one might expect, however, this primitive method of clustering fails when two materials of similar reflectivity are present in the scene (**Figure 24**)).

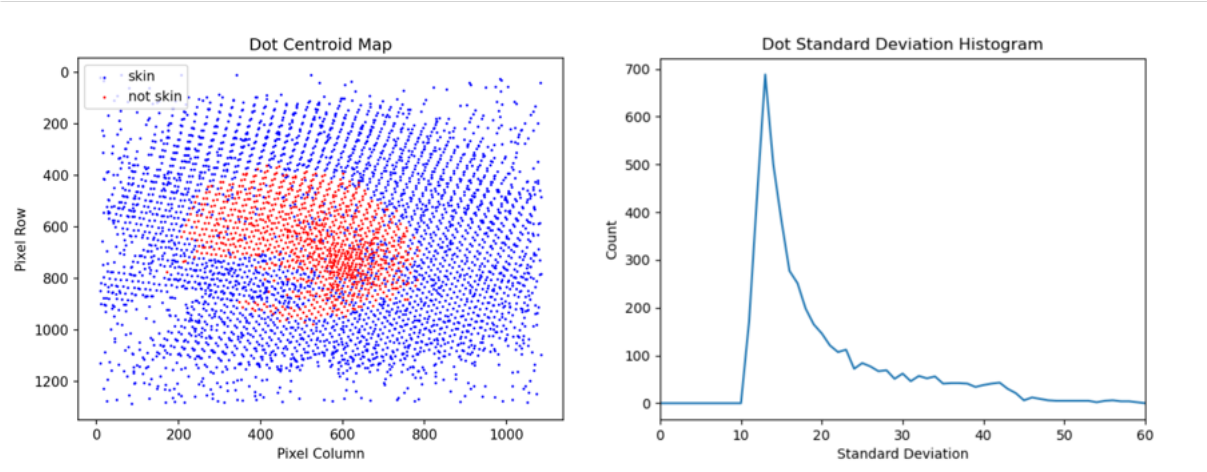

Figure 24: *Beam profile clustering using a binary threshold of the standard deviation of each beam profile for a scene including skin, metal, MDF and paper.*

As previously touched upon, these results were generated using a 2.2MPx image sensor with a dot pattern projector producing 15,000 dots. This resulted in an effective resolution of 12x12 pixels for each projected dot. In order to continue with this technique, a higher resolution of the beam profiles was required. Thus, despite promising preliminary results, investigation of this technique was closed due to limitations in available hardware.

5.2 Laser Speckle Variation Analysis

As investigated in the literature review, it is possible to use variation in the contrast of laser speckle to detect organic material. In traditional implementations of this technique, a source image whose signal is dominated by laser speckle is normally used. Paradoxically, in a DMS application, large amounts of laser speckle are undesirable due to their detrimental impact on image signal-to-noise ratio (SNR). Accordingly, the project constraints dicatated this method should be implemented with the minimum required amount of laser speckle, taking into account the high image SNR requirement of existing DMS algorithms.

In order to address this constraint, it was decided that a VCSEL illuminator with a low amount of laser speckle (and thus, a high SNR) would be used, and image processing techniques would be applied to source images to isolate the laser speckle. Theoretically, this would allow for laser speckle variation analysis to be implemented with minimal sacrifice to image SNR. However, this alone is a challenging image processing problem due to the lack of any frequency content information in laser speckle meaning frequency-based processing techniques are of little utility. Instead, the behaviour of laser speckle is much the same as noise (i.e. lacking in any 'signal' information). With this understanding in mind, a primitive method of speckle isolation was devised where a source image was first 'denoised' using a normalised Gaussian filtering operation followed by a median filtering operation. This 'denoised' image was then subtracted from the source image, theoretically leaving an image consisting only of the noise (including laser speckle) in the source image (**Figure 25**). An example of the output of this filtering operation can be seen in **Figure 26**).

To confirm the efficacy of the noise isolation function, a dataset of LED illuminated and VCSEL illuminated images were compiled to compare their respective outputs when filtered using this function. Before collecting these datasets, the output irradiance of the illuminators

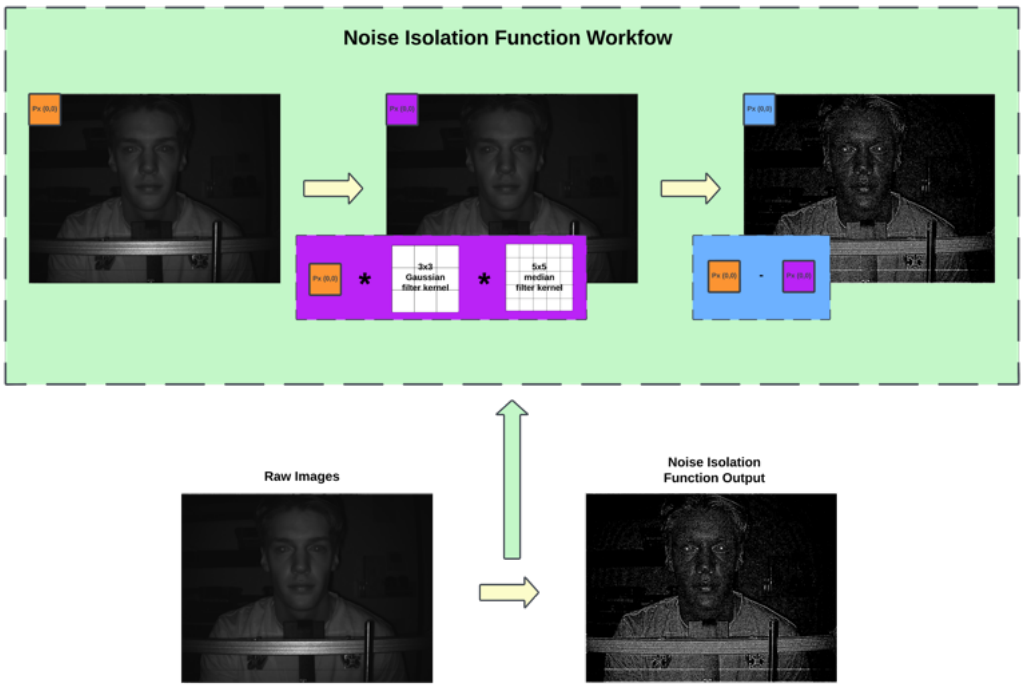

Figure 25: *Workflow of the noise isolation function.*

Figure 26: *Binary thresholded image noise obtained through subtraction of a 'denoised' image from the raw source image.*

was matched to ensure a fair comparison could be made. Then, the peak signal-to-noise ratio (PSNR) was computed across randomly sampled subsets of both of the datasets. Seeing as the behaviour of speckle is comparable to that of noise, we should expect a lower PSNR for the VCSEL images processed using the image noise isolation function. The results of the PSNR calculations supported this hypothesis, with the PSNR of the VCSEL dataset being consistently smaller than that of the LED dataset **Figure 27**).

Although these results indicated laser speckle was preserved in the output of the noise isolation function, significant non-zero output could still be seen in LED images processed by the function. This suggested that some signal (other than speckle and noise) was also preserved by the filtering operations. Insight into what this mystery signal might be can be derived from observation of the functions output. Viewing a binary thresholded output of the function, it can be seen that specular reflection highlights—such as corneal reflection and oily skin—are preserved. This suggests specular reflection is also preserved by this filtering operation. Qualitatively, this can be confirmed by observing the output of the noise isolation function on a highly reflective surface while increasing the angle of incidence **Figure 28**.

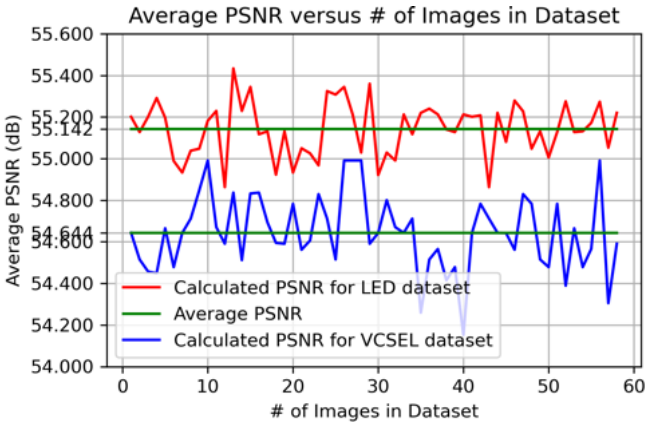

Figure 27: *PSNR comparison between output of the image noise isolation function for VCSEL and LED datasets.*

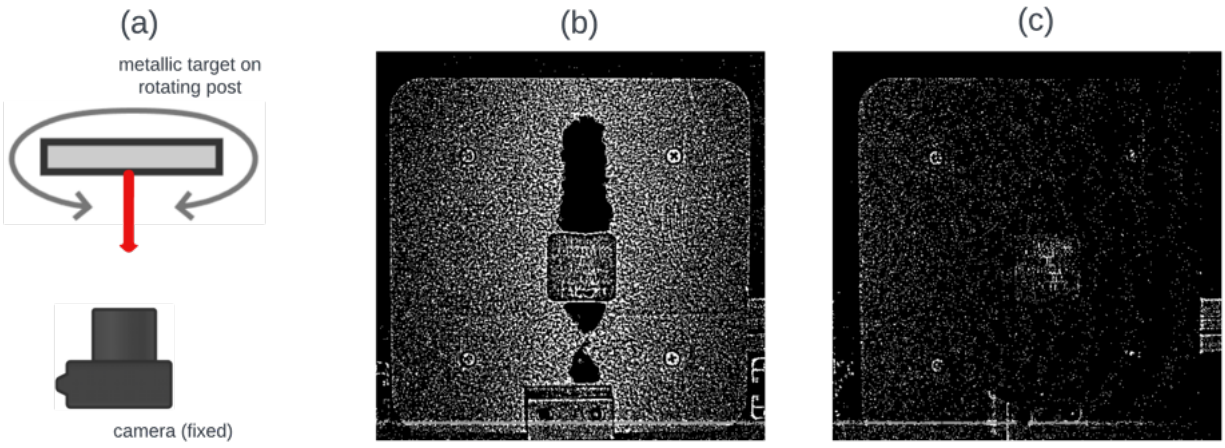

Figure 28: (a) *Experimental setup for qualitative measurement of proportion of noise isolation function output for which specular reflection is responsible. The camera uses NIR illuminators to light the scene. In the pictured setup, the angle of incidence for photons emitted by these illuminators is parallel to the surface normal (pictured in red)* [32]. (b) *Binarised output of noise isolation function for angle of incidence parallel to the surface normal. The black band observed in the middle of the target is caused by saturated (i.e. overexposed) pixels.* (c) *Binarised output of image noise isolation function for the same target with angle of incidence equal to 15°.*

Crude preliminary skin segmentation attempts using the output of the image noise function were unsuccesful. This may have been caused by a number of factors. Namely, (1) the source image may not have being 'speckly' enough, (2) the noise isolation function may have zeroed too much of the initial signal, and (3) 8-bit image quantisation may have being too low. Ultimately, it was decided that further pre-processing of the source images was required before classification could be attempted.

5.3 Laser Speckle Contrast Imaging

As established in the literature review, it is possible to use LSCI to produce vein maps. This technique could be used as a form of biometric authentication or to detect regions of bare skin. As with laser speckle variation analysis, to effectively utilise LSCI, it is important that the signal of source image(s) for which the contrast is computed is dominated by laser speckle. However, as already established, such images are of little other utility in a DMS application context due to the detrimental effects of laser speckle on image SNR.

Again, in order to address this constraint, the output of the noise isolation function described in **Section 5.2** was used as the input for the LSCI computation. Intitially, a single one of these images was used to generate a LSCI by calculating the contrast for a given pixel using local spatial (neighbourhood) information. The resulting LSCI image (pictured in **Figure 29**) contained too much noise to make any useful deduction.

Figure 29: *Example of spatially computed LSCI image.*

Next, a temporal LSCI image was computed by performing contrast calculations over a sequence of successive images. The workflow for this process is described in **Figure 30**.

This resulted in significantly less noise as demonstrated in **Figure 31**, however, the downside to this method is it requires the subject to sit still for prolonged periods of time. Note that other LSI image processing techniques presented by Dolan *et al.* were also attempted, however, they were found to produce poor outputs. These results are shown in **Appendix A**.

Although the results of the temporal LSCI computation do not show any veins, a number of interesting features do emerge. Firstly, it is immediately apparent a clear distinction arises between highly specular (i.e. oily) patches of skin and 'bare' regions of skin. Upon closer inspection, a clear distinction can also be made between 'shadowed' skin bordering the edge of the face where incident illumination is low, and also skin that is occluded by facial hair. Finally, a clear distinction can be made between different materials present in the frame such as clothing, the chinstrap upon which the face is rested, and the hard plastic block upon which the chinstrap rests, etc.. Examples for some of these categories are provided in **Figure 32**.

In summary, although the temporal LSCI LSI processing computation did not prove to be effective in fulfilling the initial intention of producing vascular mapping images, it does an excellent job at increasing the contrast between regions of different textures.

Notably, the same increase in contrast between regions of different textures is not observed when inputting raw, unprocessed images into the LSCI computation (**Figure 33**).

Surprisingly, however, no visual distinction can be made between LSCI images produced using LED illuminated source images processed by the noise isolation function versus VCSEL

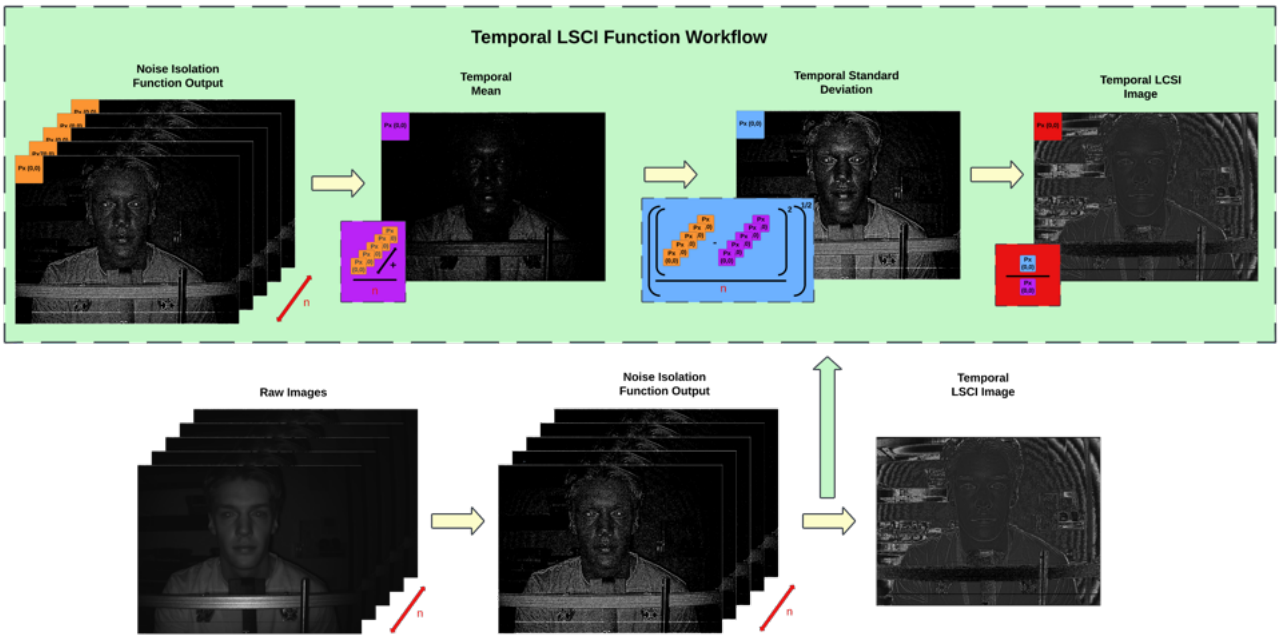

Figure 30: *Temporal LSCI function workflow.*

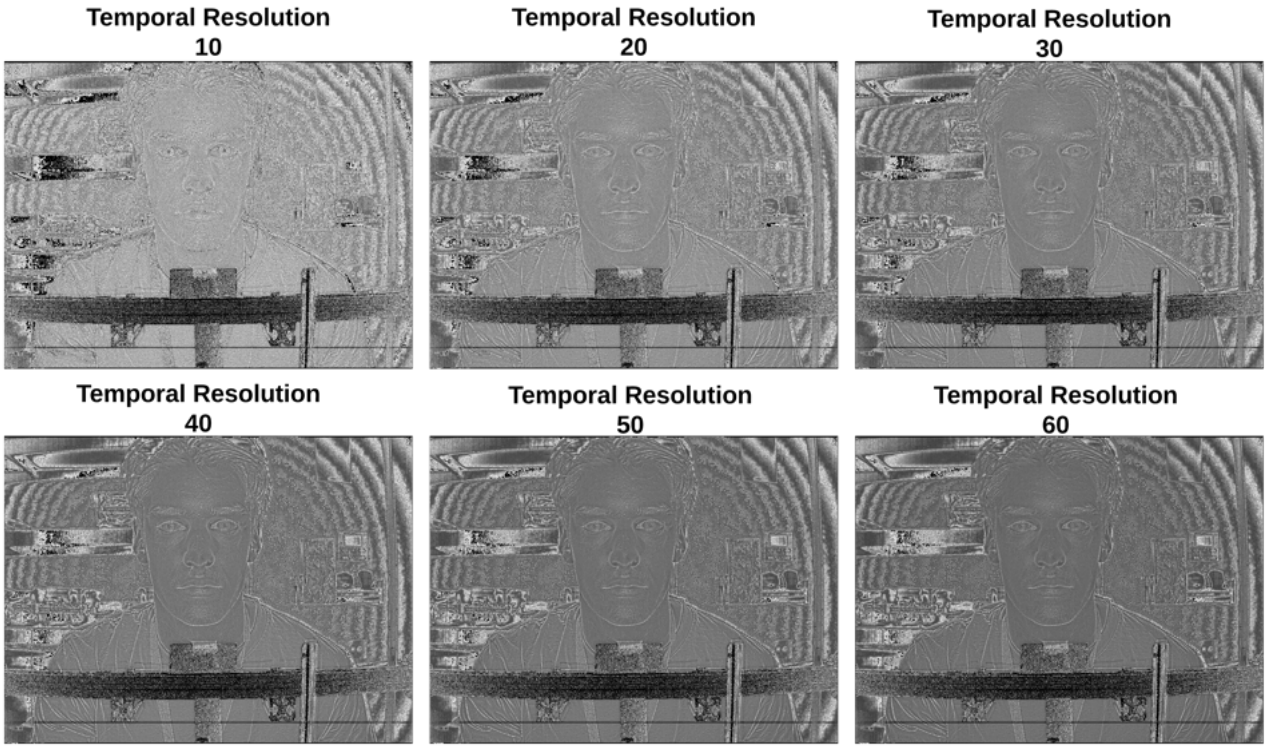

Figure 31: *Temporal LSCI images with a range of temporal resolutions. As temporal resolution increases, noise in the resulting LSCI images is reduced.*

illuminated source images. This suggests that **specular reflection within the input images is the dominant source of signal in the output of the temporal LSCI function**. Interestingly, it also suggests that, to the human eye, specular reflection is an important component in differentiating between different textures. Despite this, the background theory upon which this method is hypothesised predicts that LSCI images produced using VCSEL-illuminated

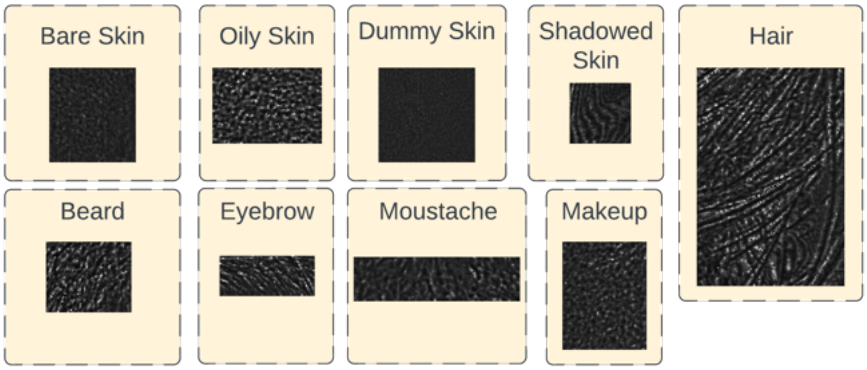

Figure 32: *Example of categories identified on the face for an LSCI image with temporal resolution 60.*

Figure 33: *Temporal LSCI images with temporal resolution 60 using a raw, unprocessed image sequence as input. In this case, the output of the LSCI operation is closer to an edge detector (rather than increasing textural contrast).*

source inputs should provide a consistent boost in classification accuracy due to the additional classification signal provided by laser speckle.

At this point, it was desired to create a classification deep-learning neural network to determine whether temporal LSCI images could be used to discern patches of 'bare' skin from a face. Seeing as equivalent increases in textural contrast were observed using LED or VCSEL illuminated source images, separate (but equivalent) datasets for each illumination source should be compiled. Classifiers may then be trained on either of the dataset and their results may be compared to ascertain insight into the importance of laser speckle in the classification process.

6 Methods

This section attempts to determine the viability of the classification method proposed in **Section 5.3** in a highly controlled setting. Such control is used to simplify the problem statement and to compile equivalent VCSEL and LED illuminated datasets. Assuming control variables are appropriately identified and maintained, a meaningful comparison can be made between analysis stemming from the VCSEL and LED datasets. Theoretically, any differences observed between classification results obtained on each dataset can then be attributed to the presence of laser speckle (i.e. the difference between the VCSEL and LED illumination characteristics).

6.1 Data Collection

Careful thought was put into the data collection process such that appropriate control variables were identified and maintained. Broadly speaking, consistent imaging, illumination and positioning between each compiled dataset was desired. Accordingly, the data capture process was designed in such a way that these variables could all be controlled between isolated rounds of data collection.

Control of imaging was relatively straightforward; a single camera was used for all data collection rounds with the lens cleaned before each capture. The same capture settings were maintained at all times (i.e. exposure=2300μs, gain=1). The camera was equipped with a 2.2MPx CMOS image sensor and a lens with a 50° horizontal field-of-view (FOV) and 40° vertical FOV.

Control of illumination was substantially more difficult due to the desire for both LED and VCSEL illuminated datasets. To minimise variations between the LED and VCSEL datasets, it was decided that intermittent pulsing of either illuminator was desired (i.e. LED illuminated frames interposed by VCSEL illuminated frames). Existing Seeing Machines (SM) technology allowed for intermittent pulsing of multiple LEDs, however, these LED drivers were not capable of producing sufficient current to drive a VCSEL. Thus, a custom method for pulsing of the VCSEL was designed.

In the first iteration of this design, a 555 timer a-stable multivibrator circuit was used to control pulses from an electronic load that drove the VCSEL with a duty cycle corresponding to the frame rate of the camera (**Figure 34**).

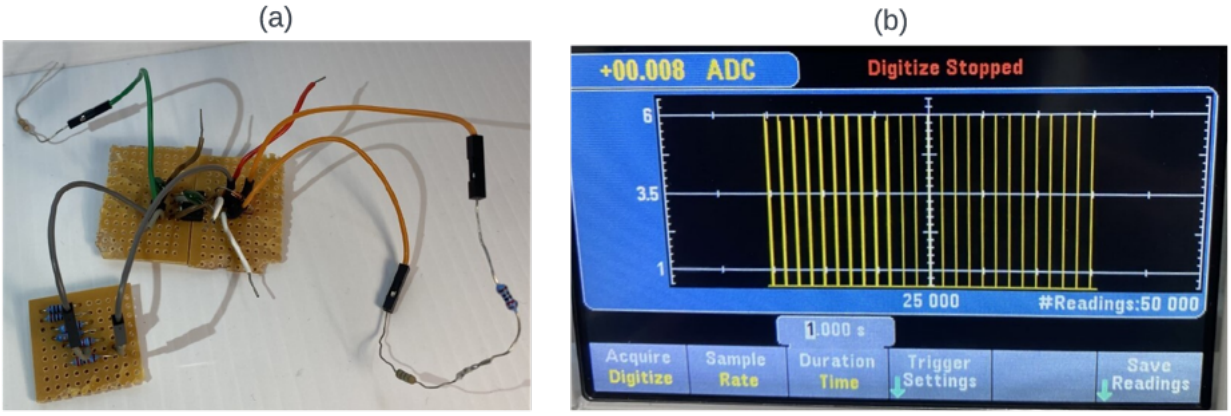

Figure 34: (a) *25Hz 555 A-stable multivibrator circuit with 1% duty cycle.* (b) *Resultant recorded current waveform when used to trigger output of electronic load.*

Although this was effective at pulsing the VCSEL, dropped frames meant it was impossible

to sync pulses from the timer with the exposure of the camera resulting in extremely inconsistent lighting conditions between sequential VCSEL frames. To overcome this deficiency, a wire was soldered to the logic signal from the LED driver board and used to trigger pulses from an electronic load, driving the VCSEL in sync with the exposure of the camera (**Figure 35**).

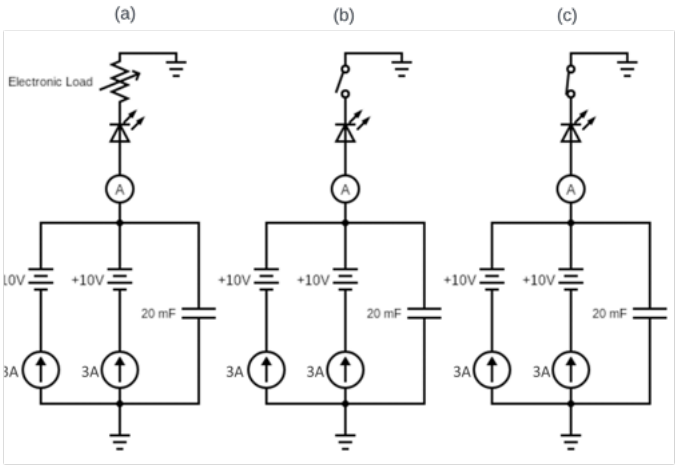

Figure 35: (a) *VCSEL power circuit diagram.* (b) *The electronic load acts as an open circuit when the LED drivers logic signal is low.* (c) *The electronic load acts as a closed circuit when the LED drivers logic signal is high. The resultant output is a square waveform pulsing the VCSEL as synchronised by the LED driver.*

With pulsing of the VCSEL synced with the exposure of the camera, it was now possible to capture frames intermittently illuminated by LED and VCSEL light sources. Remaining steps taken to ensure consistent lighting conditions between sequential frames included positioning of the VCSELs and LEDs as closely together as possible (**Figure 36 (a)**), matching of their respective irradiance's using a photometer and an analog-to-digital converter (**Figure 36 (b)**), and shielding the data capture environment from any ambient light or other undesired light sources.

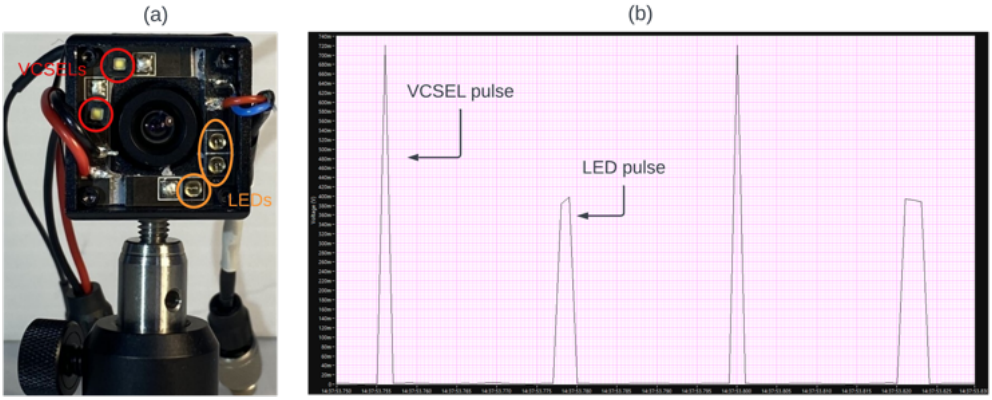

Figure 36: (a) *Surface mount soldered VCSELs and LEDs mounted to data collection camera.* (b) *Output of photometer read through analog-to-digital converter and used to match the relative irradiance of the VCSEL and LED.*

Finally, a capture rig was constructed out of aluminium extrusion to ensure consistent

positioning of subjects relative to the camera between data collection rounds and sequential frames (**Figure 37**).

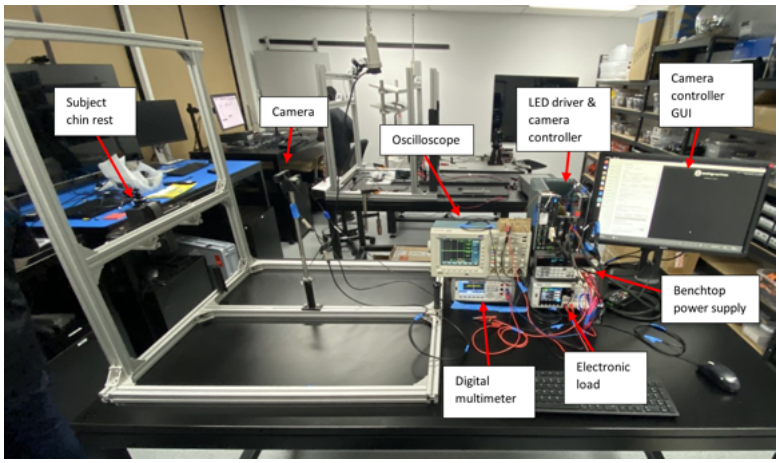

Figure 37: *The completed data capture rig.*

With the implementation of these control mechanisms, the data capture process was deemed sufficiently comprehensive for data collection to begin. A total of three rounds of data collection were completed. These provided approximately 84,000 images across 23 different subjects.

6.2 Image Processing

Images compiled in **Section 6.1** were processed as described in **Section 5.3**. For each subject, LED and VCSEL frames were separated and saved in groups of 60 consecutive frames. For each of these groupings, corresponding LSCI images were computed with temporal resolutions 5 to 60 (inclusive) in increments of 5. For any given temporal resolution and illumination source, a total of 1400 LSCI images were produced. Thus, across each temporal resolution and for both illumination sources, 33,600 unique LSCI images were produced.

6.3 Image Labelling

A manual masking tool was created to facilitate labelling of the collated datasets. Attempts were made to automate this process with an unsupervised texture segmentation algorithm (see **Appendix B** for more details), however, it was quickly realised that optimisation of this algorithm for the speed required to annotate all 1400 LSCI images in a timely manner was too time consuming for a tangential task. Comparatively, the manual masking tool took only two days to create, and was able to increase productivity to where all collected data was masked in a single day. The basic workflow of the manual masking tool is summarised in **Figure 38**.

The masking tool takes a filepath to a folder of LSCI images as the raw input. When a user enters a filepath, 'block one' is used to extract masked ROIs from the first image. If a user decides to save a mask, keybindings are used to determine the label given to the mask. All subsequent LSCI images are passed through 'block two'. Assuming the subject stayed still for the duration of the data capture process, block two allows the image labelling workflow process to be greatly expedited.

Each LSCI image was stored in a folder containing a source pre-processed image that could be toggled to overlay the LSCI image if the user desires to resolve any detail that is unclear in the LSCI image. Masked ROIs made on an LSCI image were stored within the same folder and the coordinates and labels of each masked ROI were stored in a comma-separated variable (CSV) file.

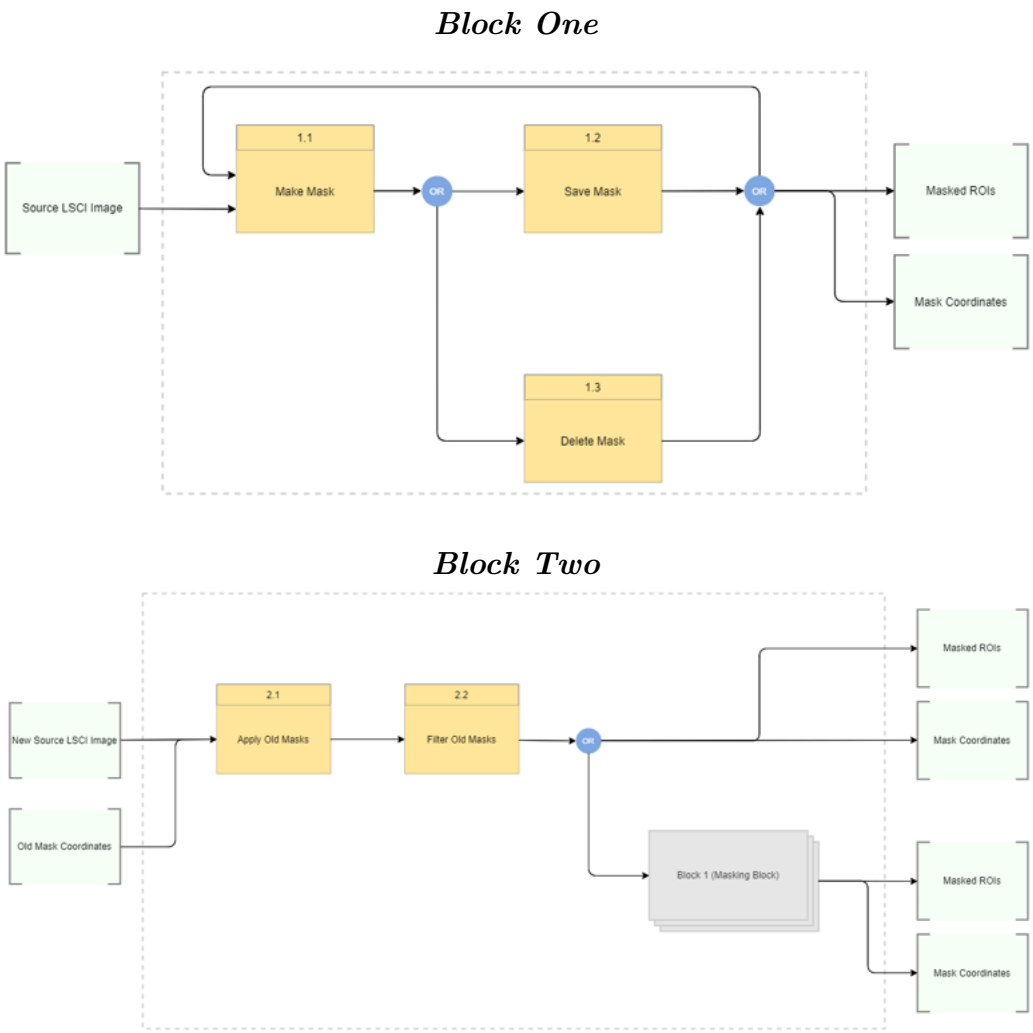

Figure 38: *Functional flow block diagram summarising the masking tool workflow. Block one encapsulates the basic masking process. Block two illustrates the application of previously made masks to the next source LSCI image (where the filtering block let's you individually choose whether to save or delete these masks), as well as the opportunity to apply new masks after having filtered the old masks.*

Using the masking tool, 9,791 individual masked ROIs were created with 3,075 of these being of 'bare skin'. The remaining, 6,716 images include categories: oily skin, shadowed skin, hair, facial hair and makeup. Masks were made on the VCSEL dataset with temporal resolution 60 as differences in textural contrast were most easily distinguishable in this dataset. However, the CSV file containing the label and coordinate information for each mask meant that the same masked ROIs could be applied to both the VCSEL and LED illuminated datasets and to equivalent datasets using lower temporal resolution LSCI images (where the textural differences were not as easy to distinguish). For example, given range of 120 consecutive frames, these frames were split into a VCSEL and LED dataset (i.e. 60 VCSEL frames and 60 LED frames) and a temporal LSCI image was produced for both datasets. Then, assuming a mask of the left cheek is made on the VCSEL LSCI image, the CSV file was used to apply the same mask to the LED LSCI image. This meant that two equivalent masks were created, with the only difference being that one was based on VCSEL-illuminated data and the other was based on LED-illuminated data. Additionally, if a fraction of the 60 images was sampled to create an

LSCI image with lower temporal resolution, the CSV file could again be used to get equivalent masks, with the only difference being the lower temporal resolution of the source LSCI image. Using this process, 24 unique datasets were created each consisting of recordings from all 23 subjects. Each of these datasets contained the same 9,791 masked ROIs, meaning a total of 234,984 unique masked ROIs were created. These labelled ROIs formed the datasets on which different classification models were trained.

6.4 Supervised Learning

Supervised learning refers to machine learning using labelled training data. In comparison to unsupervised learning, supervised learning is commonly associated with a higher degree of accuracy and greater control over the training process. There are many ways supervised learning can be approached. Plested *et al.* details some common supervised learning techniques used in image classification [33]. One such method is transfer learning, which this project implements using a ConvNeXt convolutional neural network (CNN) as the backbone architecture.

For computer vision problems, CNNs have long being the gold standard of accuracy [34]. This is generally not seen as a coincidence, as the functioning of a traditional CNN draws many parallels to that of the human visual cortex. Furthermore, CNNs have built-in inductive biases (such as translational equivariance) that are desirable properties for vision-based classification tasks.

The ConvNeXt model was fine-tuned on all of the datasets compiled in **Section 6.3**, however, the performance evaluation of accuracy versus temporal resolution (**Section 7.1**) found accuracy to plateua above temporal resolution 25. Seeing as the minimum required temporal resolution is desired in a DMS application context, the analysis presented in this section is only for models trained on temporal resolution 25.

6.4.1 ConvNeXt design philosophy

In the year 2020, the introduction of Vision Transformers (ViT) reformed the field of computer vision, achieving state-of-the-art accuracy on the ImageNet dataset [35]. In a traditional computer vision problem, this result was surprising, seeing as transformers draw far less parallels to the human visual cortex than CNNs, and were initially confined to natural language processing problems. Following this development, many believed transformer-based networks and their unmatched computational scalability poised them to be the dominant generic vision model backbone. However, ViT's global attention design means it has quadratic complexity with respect to input size [36]. This becomes inefficient when higher-resolution inputs are passed (versus the standard 224x224 pixels use in ImageNet).

Heirarchical transformers attempted to address this weakness by introducing attention within local windows. Theoretically, this would combine the most desirable traits of transformers and CNNs: scalability and local attention, respectively. However, rather ironically, these amendments can be viewed as a naive attempt to implement convolution in transformer networks at the cost of increased design complexity and/or reduced efficiency [34]. This observation lead to the question: *if the performance of transformers in computer vision problems can be improved by increasing their similarity to CNNs, can the performance of CNNs be improved by amending their architecture to be more similar to transformers?*

In 2022, Zhuang Liu *et al.* sought to answer this question by taking a standard ResNet and gradually modernising the training procedure and architecture by drawing inspiration from hierarchical vision transformers. The result of this paper is the ConvNeXt family of CNN architectures, which were found to compete favourably with transformers in terms of accuracy, scalability and robustness across all major benchmarks [34].

The authors decompose their design amendments into two categories: macro and micro design. On the macro level, changes are made to the stage compute ratio and the "stem cell" structure. On the micro level, ReLU activation functions are replacd with GELU, fewer overall activation functions and normalisation layers are used, batch normalisation is replaced with layer normalisation and downsampling and normalisation layers are added within stages (rather than only at the start of stages). Other changes include: use of depthwise convolution, inverted bottleneck stage restructuring (i.e. hidden dimension is wider than the input dimension), and larger kernel sizes (to mimic global attention). With these design amendments, the resulting model achieves an accuracy of 82% on ImageNet, versus Swin-T's 81.3%. Notably, all design amendments are adapted from vision transformers.

To compare scalability, the authors create larger ConvNeXt models. In all cases, ConvNeXt models achieve higher accuracy scores and faster throughput than Swin Transformers with a similar number of parameters.

The findings of Zhuang Liu *et al.* challenge the widespread belief that the importance of convolution in computer vision problems is diminishing. They also suggest that nuanced design consideration remains important in machine learning, rather than simply maximising the number of model parameters. In many ways, the ConvNeXt architecture is to CNN's what hierarchical transformers are to transformers.

6.4.2 ConvNeXt model architecture

Section 6.4.1 presents a high-level understanding of the design philosophy with which ConvNeXt was created. A deeper understanding of the model can be attained by examining the input-to-output model datastream.

Excluding the stem stage, the ConvNeXt model consists of four sequential stages. Majority of these stages contain an initial depthwise convolution operation with stride two that downsamples the input while simultaneously doubling the numbers of extracted features. To stabilise training, the downsampling operation is preceded by a layer normalisation (without this, training diverges). Following the downsampling operation, each stage consists of a ConvNeXt block (3:3:27:3 respectively mimicking the the 1:1:9:1 stage compute ratio used by Swin-T) that contains an initial depthwise convolution feature extractor (with a large 7x7 kernel size mimicking global attention of transformers) proceeded by a layer normalisation operation (to prevent internal covariate shift) followed by a multilayer-perceptron (MLP) with an inverted bottleneck design (such that the hidden layer of the MLP is four times wider than the input dimension). This internal structure is summarised in **Figure 39 (a)** and **Appendix C**.

6.4.3 Implementation setup

In this method, a CNN classifier using a ConvNext backbone is created to differentiate bare skin from other categories observed in temporal LSCI images (**Figure 32**). Given an input of bare skin, the model should return a prediction of 1 (positive). For other inputs, the model should return a prediction of 0 (negative). Some examples inputs and corresponding predictions produced by the model can be seen in **Figure 38 (b)**.

Before fine-tuning the model to the use-case, it is important to define exactly what it is hoped the model will learn. This allows for identification of any undesirable features the model might instead learn so the training data may be augmented/manipulated such that this is avoided. In this use-case, it is desired that the model learns to identify skin based off it's textural appearance in the LSCI images produced in **Section 5.3**. Specifically, it is desired that the model learns how to distinguish bare skin from other categories such as: oily skin, shadowed skin; and skin that is occluded by facial hair, hair or makeup. Ultimately, it is

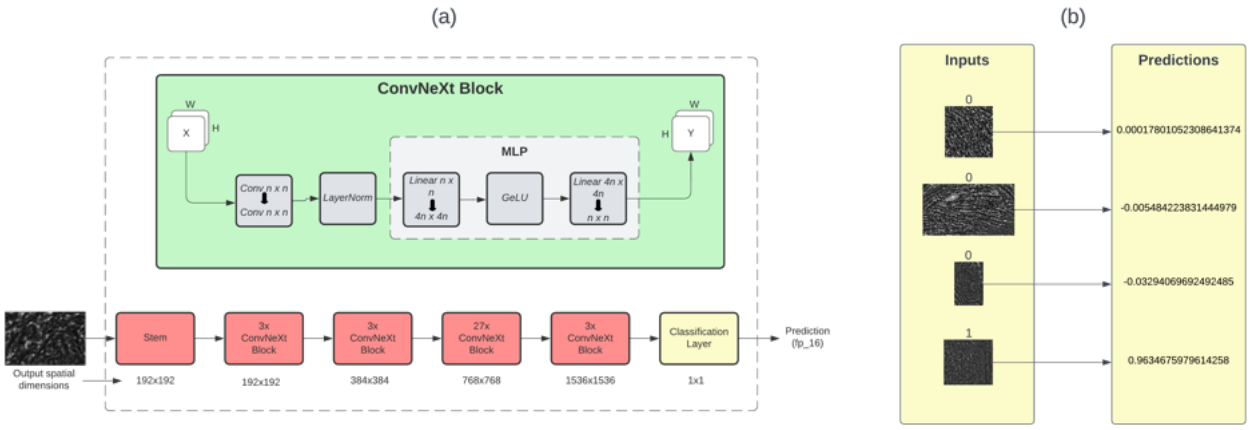

Figure 39: (a) *Summary of the model architecture used for classification in this project. A supervised learning approach was used where the model was fine-tuned using a ConvNeXt backbone (pictured in red).* (b) *Examples of inputs and corresponding model predictions.*

desired that the skin classifier can be implemented with a pre-existing facetracker such that classifications may be made on different regions of the face.

Although it is expected that classification will be possible using LSCI images constructed from both LED or VCSEL illuminated source inputs, the theory on which this method was justified hypothesises that VCSEL illuminated source inputs should provide a consistent boost in classification performance due to the additional classification signal provided by laser speckle. The datasets compiled in **Section 6.3** will be used to test this hypothesis.

6.4.4 Model Training

Dataset details

When training a supervised deep learning network, it is important to appropriately partition the labelled data to prevent overfitting. Accordingly, 10% of each category from the labelled datasets compiled in **Section 6.3** were extracted to be used as a test dataset. The remaining data was 80/20 split into a training and validation set.

Training setup

For fine-tuning, the training data was fed into the model in batch sizes of 16, with gradient accumulation used to update the model parameters (i.e. weights) every four batches. This meant the model could be trained on a workstation with a single NVIDIA RTX A4000 taking one hour. The model was fine-tuned for 12 epochs. More epochs were not used as training and validation losses had already plateuad (**Figure 40**).

Dataloader setup

When creating the models dataloader, the training and validation data was zero-padded or 'squished' to resolution 224x224 pixels. This was done as the ConvNeXt backbone architecture was trained on ImageNet, meaning it was optimised for 224x224 pixels inputs. Before a forward pass through the model, each batch was augmented with the transformations described in **Table 2**. An example of a batch processed by the dataloader can be seen in **Figure 41**. These augmentations were derived from an ablation analysis and the respective probability of appli-

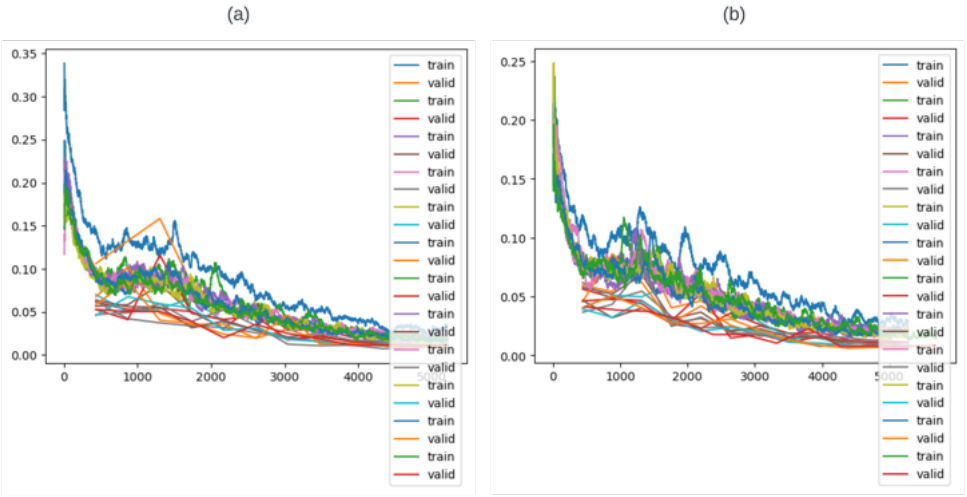

Figure 40: (a) *Training and validation losses for the LED models trained on temporal resolutions 5-60.* (b) *Training and validation losses for the VCSEL models trained on temporal resolutions 5-60.*

cation for each transformation was determined through an augmentation percentage derivation script. These augmentations were applied in an effort to prevent overfitting and increase the robustness of the model.

Table 2: *Batch augmentation used for model training.*

Augmentation	Magnitude	Probability
Warp	0.2	0.3
Brightness	0.2	0.06
Flip	N/A	0.61
Contrast	0.4	0.23
Saturation	0.2	0.35

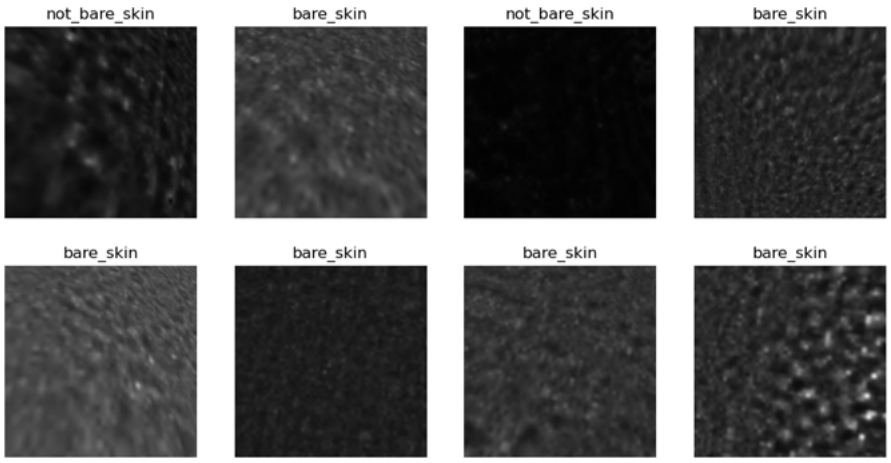

Figure 41: *An example of an augmented batch output by the models dataloader.*

Performance analysis

The regression output datablock returns a single value that is thresholded to make a prediction. A receiver operating characteristic (ROC) curve analysis was used to determine the optimal thresholding value (**Figure 42**). The red dot on the true positive rate (TPR) versus false positive rate (FPR) curve corresponds to the threshold value producing the maximum difference between TPR and FPR on the ROC curve (∼0.33).

ROC curves tend to provide overly optimistic illustrations of the model on datasets with a class imbalance. Seeing as there was a ∼1:2 ratio between the bare skin and not bare skin categories, a precision recall curve was used to examine the trade-off between precision and recall (**Figure 42**). The red dot on the precision versus recall curve corresponds to the precision versus recall score for the identified optimum ROC classification threshold (∼0.33). In the final model, this classification threshold was used as the corresponding point on the precision recall curve provided a sufficiently low false positive and negative rate for the use case.

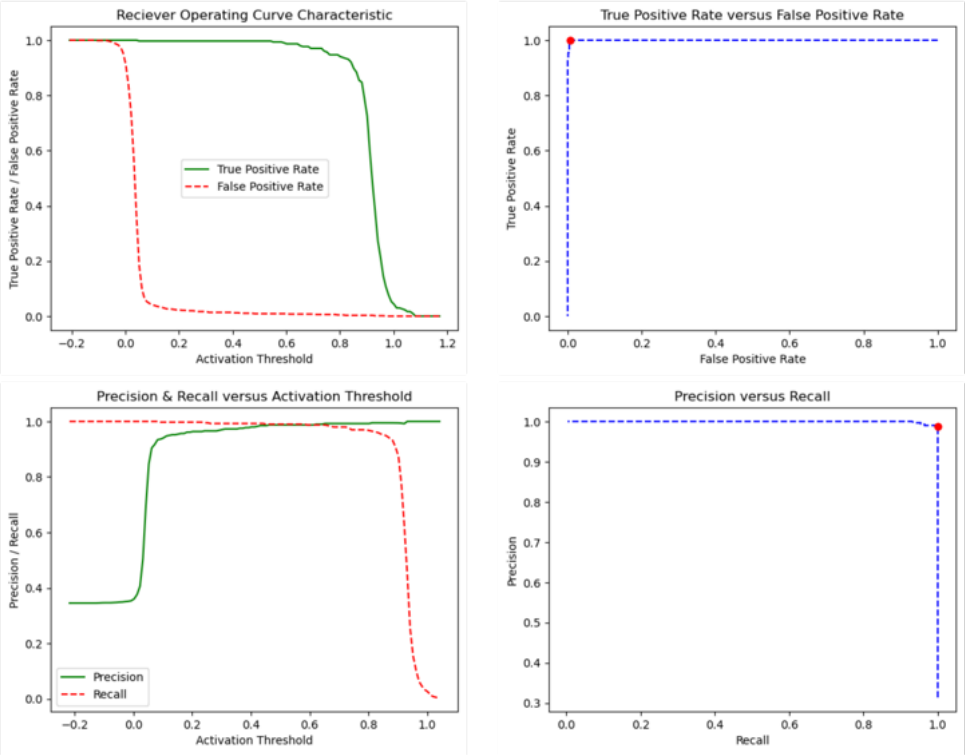

Figure 42: *ROC and precision recall analysis of the regression classification model.*

Finally, an ensemble model was created by conducting the same training processes using Swin-L, ViT-L and ConvNeXt-L backbones. The accuracy of the ensemble was tested using a weighted sum of predictions, mean prediction and absolute maximum prediction, however, there was no noticeable improvement in accuracy performance versus a single ConvNeXt-L model. Accordingly, the ensemble model was not used in the final evaluation.

As an aside, a categorical classification model was also created that returns predictions for all categories in the training data - i.e. bare skin, oily skin, shadowed skin, facial hair, hair, makeup and dummy skin (aka. fake skin) - for a single input. This model achieved an impressive accuracy of 94.525%, however, was not investigated further due to scoping limitations. More information on this model is included in **Appendix D**.

6.5 Performance Evaluation

In this section, the performance of the final model on the test dataset is presented and attempts are made to provide insight into what the model has actually learned. A confusion matrix summarising the performance of the final model on the test dataset is shown in **Figure 43**.

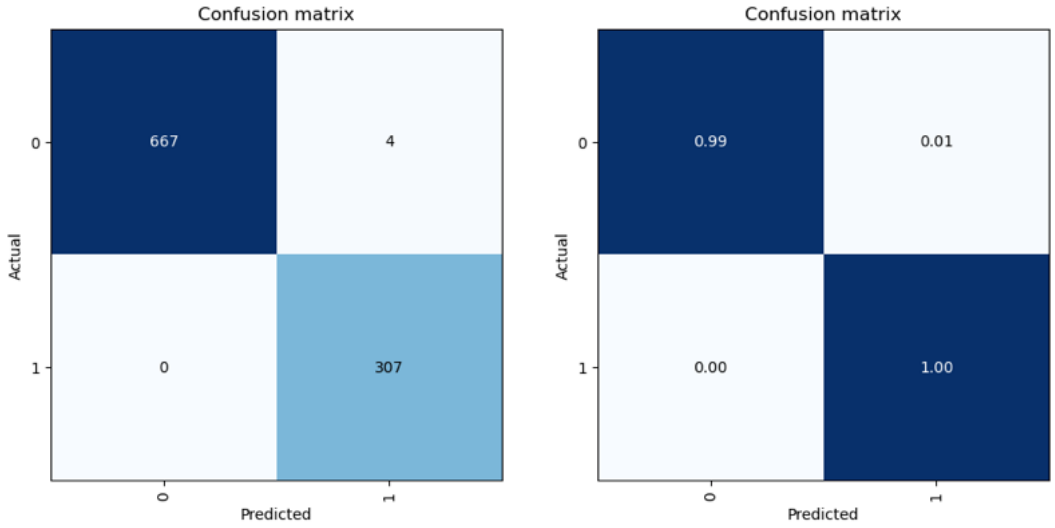

Figure 43: *(left) Absolute and (right) relative confusion matrix summarising performance of the VCSEL model on the VCSEL test dataset.*

The model performs well on the test dataset, suggesting the problem statement is solveable using the given technique. Furthermore, in line with the experimental hypothesis, the VCSEL model outperformed the LED classification model, although only by a small margin (∼0.2%). Because of the careful thought spent on maintaining control variables between the two test datasets, it can be assumed that this difference is attributable to the additional classification signal provide by laser speckle. However, it is important to note that these findings do not provide any definitive insight into whether this model could serve as an effective skin classification tool in an in-field DMS application. Instead, the findings demonstrate that it is possible to create an accurate binary skin classification algorithm based on the assumptions used in the data collection process - i.e. maintaining controlled ambient lighting and illumination conditions, consistent subject positioning, etc.

In order to make any definitive conclusions about the true viability of the technique, a rigorous evaluation process needs to be devised that systematically assesses the impact of each of these assumptions on the performance of the classification model. Additionally, any edge-cases that are not covered by the training dataset should be identified such that insight may be gathered into the robustness of the algorithm. **Section 7** attempts to produce this comprehensive evaluation.

6.5.1 Insights into Model

One powerful method of ascertaining insight into what a CNN has learned is to examine some of the feature extractor kernels. Kernels higher in the module are relatively simple and serve to extract low-level features, while kernels deeper in the module are tuned to identify, complex, high-level features. For example, **Figure 44** contains some of the kernels contained in the first ConvNeXt block of stage zero.

Comparatively, it can be seen that kernels in the last stage of the model architecture serve to extract far more complex features (**Figure 45**).

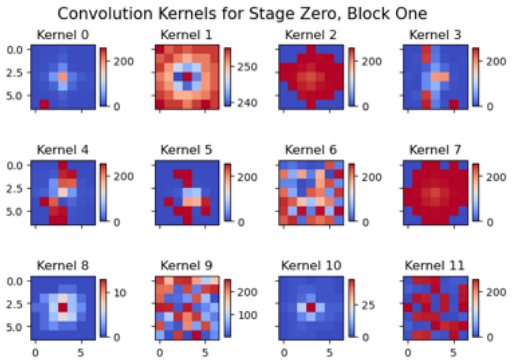

Figure 44: *A selection of the highest level feature extraction kernels in the VCSEL model from the first ConvNeXt block of stage zero. These kernels are tuned to identify simple, low-level features.*

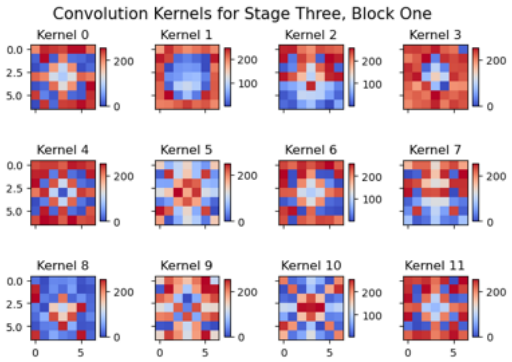

Figure 45: *A selection of the lowest level feature extraction kernels in the VCSEL model from the first ConvNeXt block of stage three. These kernels are tuned to identify complex, high-level features.*

Further insight can be provided into the learning patterns of the model by comparison of the models trained on the LED versus VCSEL datasets. **Figure 46** shows the accuracy of both models on both datasets.

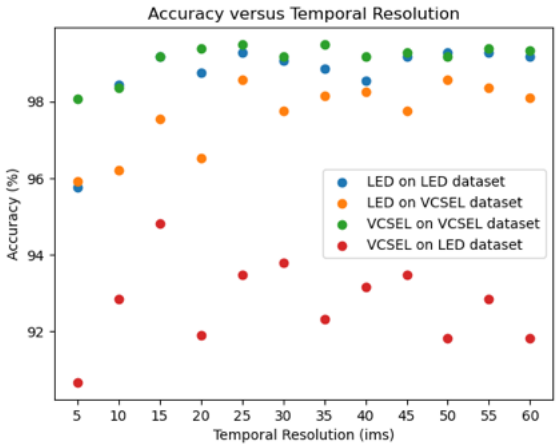

Figure 46: *Accuracy of the binary regression classification models versus resolution on both the LED and VCSEL datasets.*

Figure 46 shows that VCSEL-trained models consistently scores poorly on the LED dataset. This suggests the VCSEL model learned to classify based of a 'signal' not present within the LED dataset. This 'signal' is hypothesised to be laser speckle. Comparatively, the LED-trained model demonstrated a slight degradation in performance when evaluated on the VCSEL dataset (<2%). This suggests that the LED-trained models learned to classify off a signal present in both the VCSEL and LED datasets. It is hypothesised that the small degradation in performance observed can be attributed to the presence of laser speckle in the VCSEL dataset (which the LED-trained model would consider analogous to image noise).

Principal component analysis (PCA) can be used to better understand these discrepancies. PCA is a statistical analysis technique whereby the dimensionality of a dataset is reduced to a subset of eigenfunctions that convey the most variance in the dataset. Accordingly, some of the variance in the original dataset is lost when conducting PCA, however, enough variance generally remains to effectively visualise the clustering performed by the model. In this case, the output of both the VCSEL and LED models is clipped to the second last linear function within the classification layer. This function takes 3072 in-features and compresses them to 512 out-features. Accordingly, for a single input, the model outputs 512 individual floats at this point. PCA is used to reduce this dimensionality from 512 to 2, and the resultant clustering performed by each model on both of the test datasets is observed in **Figure 47** and **Figure 48**.

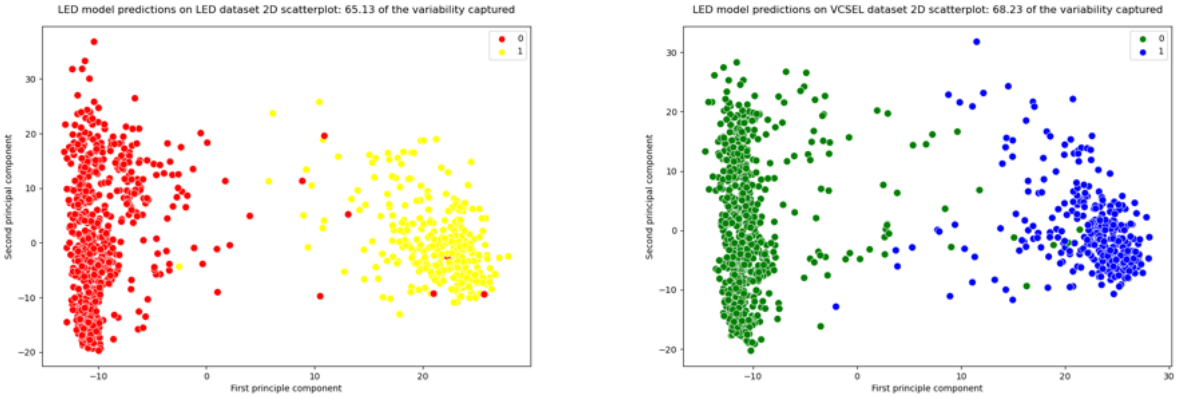

Figure 47: *LED model PCA analysis on the (left) LED and (right) VCSEL datasets.*

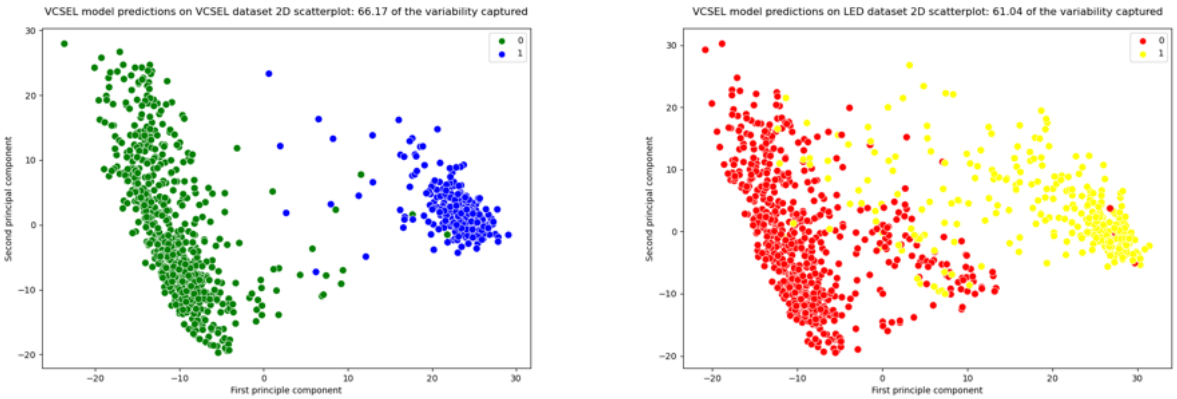

Figure 48: *VCSEL model PCA analysis on the (left) VCSEL and (right) LED datasets.*

The PCA findings indicate that, from the perspective of the LED model, there is little difference between the VCSEL and LED datasets. Comparatively, from the perspective of the VCSEL model, there is substantial difference between these datasets. These findings are in agreement with the above analysis of the accuracy results presented in **Figure 46**.

As an aside, models were also trained using masks extracted from raw frames as inputs (i.e. no pre-processing applied to the inputs). In this case, a perfect classification score is achieved by both models on both datasets. While this finding is very interesting, it is very difficult to infer what exactly the model learned in this case. This result is expanded upon in **Appendix E**.

7 Results

As aforementioned, the performance evaluation of the trained models presented in **Section 6.5** does not provide any insight into the in-field viability of the models. For such insight to be attained, assumptions baked into the training, validation and test datasets must be identified and the performance of the model must be quantified against datasets free from these assumptions. This section seeks to conduct this evaluation by systematically identifying these assumptions and compiling test datasets that can be used to estimate the impact of each individual assumption.

7.1 Performance versus Temporal Resolution

As established in **Section 5.3**, the noise in a temporal LSCI image diminishes as the temporal resolution (i.e the number of images over which LSCI is computed) increases. In a DMS application context, it is desirable to use the minimum temporal resolution possible. This is because higher temporal resolutions have increased processing requirements and are more susceptible to motion blur. While a temporal resolution of 60 allowed for easy manual labelling of masked ROIs, textural differences remained apparent in LSCI images with much lower temporal resolutions. Accordingly, it is likely a classification algorithm could be trained to make classifications on LSCI images with temporal resolution less than 60. In this section, such a possibility is investigated by retraining models using LSCI images with temporal resolutions ranging from 5 to 60. To ensure a fair comparison, it was important the same masked ROIs were used in the training, validation and test datasets for all models. As explained in **Section 6.3**, this meant that all models were trained and validated on the same data, with the only difference being the temporal resolution of LSCI images in each dataset (e.g. **Figure 49**). The results of this investigation are shown in **Figure 50**.

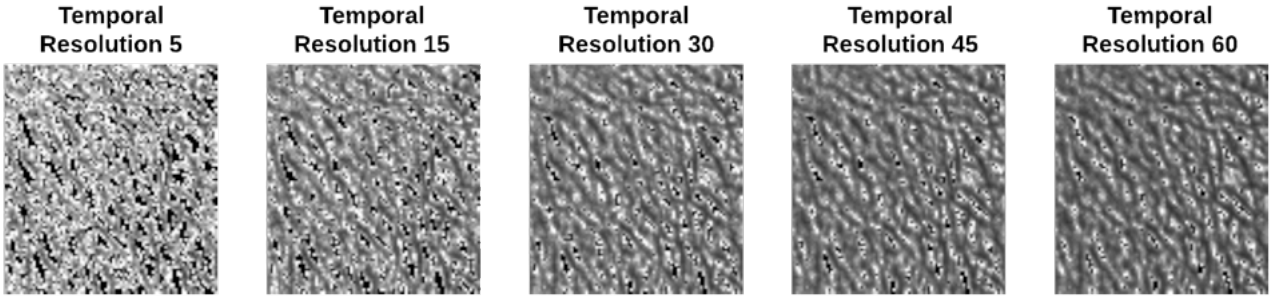

Figure 49: *An example of the same masked ROI taken from images with varying temporal resolution.*

The findings presented in **Figure 50** indicate that prediction accuracy plateaus for temporal resolution greater than 25. In fact, accuracy actually declines from this point (assumingly due to the blur associated with operating over such a large period of time). Hence, subsequent performance evaluations in this section only focus on the models and datasets associated with temporal resolution 25. Note that all previous results presented in Section 6.4 and Section 6.5 also use the temporal resolution 25 models and datasets (unless otherwise indicated).

Notably, VCSEL models consistently outperform the accuracy of LED models. This is in line with the theory upon which the logic behind this classification method was formulated.

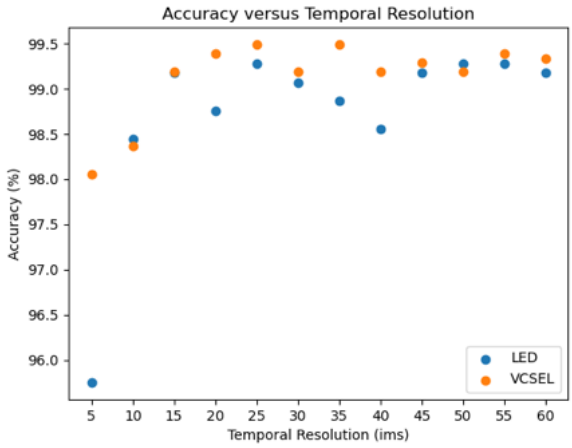

Figure 50: *Accuracy versus temporal resolution for the VCSEL and LED models.*

7.2 Performance versus Image Resolution

As discussed in **Section 6.1** a data capture rig was used for the data collection process to maintain a constant subject positioning for the duration of the capture process. While this assumption greatly simplified the masking and labelling of collected data, it meant that a constant resolution was used for the training data (∼600x400 pixels for a subjects face). In a real world application context, resolution is likely to vary as a subjects positioning changes relative to the imaging plane. Accordingly, a dataset needs to be developed that investigates the effect of image resolution on the classification performance of the model.

Such a dataset was compiled by taking the original masked and labelled images recorded using the data capture rig and downsampling each mask using Lanczos downsampling. Ignoring any inaccuracies associated with Lanczos downsampling, this meant the only difference between the downsampled datasets was the resolution of each mask. It also meant that any difficulties associated with correctly labelling masks taken on lower resolution images were avoided.

The performance of the model on these downsampled datasets is summarised in **Figure 51**. Note that the image resolution downsampling factor refers the multiplicative coefficient by which the downsampled images dimensions were computed. For example, an image resolution downsampling factor of 0.1 means a 10x10 pixels image would be downsampled to 1x1 pixels.

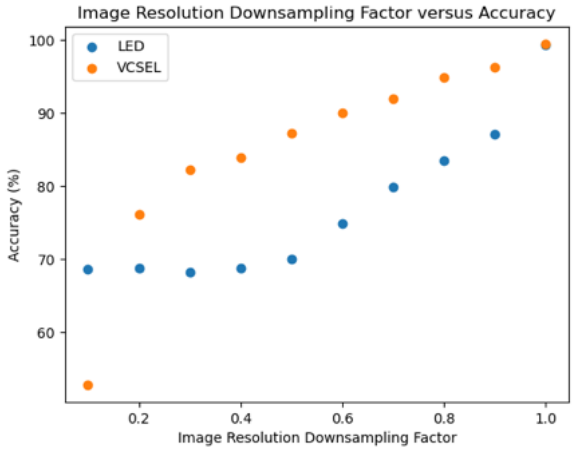

Figure 51: *Accuracy versus image resolution downsampling factor for the VCSEL and LED models.*

A simple comparison of accuracy versus image resolution can be deceptive due to the class imbalance present in the test dataset. Using confusion matrix statistics to compute the ROC area under the curve (ROC AUC) score provides deeper insight into the behaviour of the model at lower image resolutions (**Figure 52**).

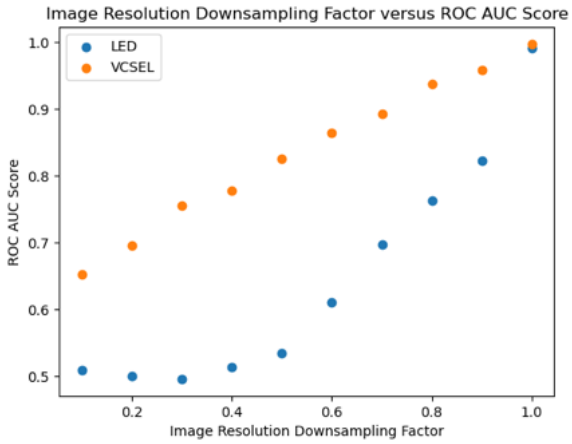

Figure 52: *ROC AUC score versus image resolution downsampling factor for the VCSEL and LED models.*

The ROC AUC analysis indicates that, in the case of the LED model, the model is effectively making random predictions for downsampling factors less than or equal to 0.5 (i.e. <= half of the original mask resolution). Contrastingly, the VCSEL model continues to produce some form of meaningful classification (albeit, a weak one) for all image resolutions. Examining the raw confusion matrix statistics provides further insight into these results (**Table 3** and **Table 4**).

Table 3: *VCSEL model confusion matrix statistics for test downsampled dataset.*

Downsampling Factor	True Positive	True Negative	False Positive	False Negative
0.1	303	213	458	4
0.2	159	586	85	148
0.3	177	627	44	130
0.4	188	632	39	119
0.5	215	638	33	92
0.6	235	646	25	72
0.7	252	647	24	55
0.8	278	650	21	29
0.9	290	651	20	17
1.0	307	667 4	0	

Table 3 indicates that the LED model becomes virtually entirely biased towards negative (i.e. 'not bare skin') predictions for downsampling factors less than or equal to 0.5. This explains the plateau that can be seen for the LED model in **Figure 51**. Comparatively, the VCSEL model sees a steady, linear decline in prediction accuracy until image downsampling factor 0.1, where predictions suddenly saturate towards positive classifications (i.e. 'bare skin'). These results suggest that the VCSEL model is significantly more resilient to decreases in image resolution than the LED model.

Table 4: *LED model confusion matrix statistics for test downsampled dataset.*

Downsampling Factor	True Positive	True Negative	False Positive	False Negative
0.1	11	651	12	291
0.2	1	662	1	301
0.3	0	658	5	302
0.4	16	647	16	286
0.5	29	646	17	273
0.6	73	649	14	229
0.7	128	643	20	174
0.8	172	634	29	130
0.9	209	631	32	93
1.0	301	657	6	1

Notably, training new models on these downsampled datasets saw significant improvements in accuracy. In this case, the VCSEL model was able to maintain >98% classification accuracy for image downsampling factors 0.2-1.0, and 95% for a image downsampling factor of 0.1. This suggests that it is possible to use the same classification technique at further distances with a more comprehensive training and validation dataset.

It is important to note that this section is distinct from 'performance versus spatial distance'. While resolution is inversely proportional to distance, many other image characteristics also change as spatial distance is increased (e.g. increased blur, decreased irradiance, etc.). Noting this, this section could be used to estimate the impact of increasing spatial distance on the models classification performance but a direct correlation cannot be made between the two!

7.3 Performance versus Ambient Light

To evaluate the overall viability of the model, it was important testing was conducted replicating illumination conditions indicative of a real DMS application. To accomplish this, a mobile workstation was created such that data capture could be performed outdoors in the cockpit of a car.

The camera module was mounted on the dash of the car in line with the center of the drivers seat, roughly 70cm from the drivers face. A photometer was used to record an average value for ambient light intensity during a recording. Additionally, variance in ambient light intensity was recorded. If the variance in ambient light over a single recording was too high (>500μW), the recording was not used in the evaluation process. For each recording, the mean value of ambient light was used to organise the recording into a qualitative category (**Figure 53**).

It is important to note that the recorded irradiance of ambient sunlight can vary substantially depending on orientation of the measuring instrument relative to the position of the sun. Consistent positioning of the photometer was used for outdoor data collection to mitigate this issue, however, little attention should still be given to the quantitative ambient light values associated with each qualitative category. Instead, attention should be focused on what the example scenes look like for each category. Unfortunately, a hardware issue with the outdoor data collection rig meant it was not possible to record LED illuminated images for this task.

Each recording was performed around the same time of day (within an hour), such that consistent positioning of the sun was maintained between each recording. The classification results for each category are presented in **Figure 54**.

As expected, the results indicate that the model performs well in conditions with low ambient

Figure 53: *Ambient light categories used for categorising of the ambient light evaluation datasets. For each category, an example of a corresponding image is provided.*

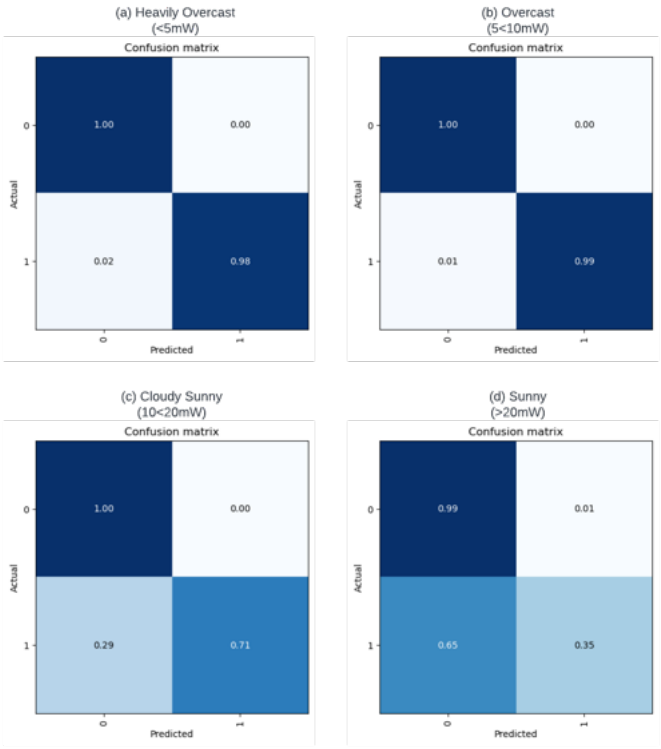

Figure 54: *Comparison of relative results for each ambient light category. A steady decline in performance can be observed as ambient light intensity increases above 10mW.*

light intensity. However, for the 'cloudy sunny' and 'sunny' categories, significant declines in classification performance are observed. Notably, the impact of ambient light on classification performance is similar to that identified for salt-and-pepper, Gaussian and Rayleigh noise of

high magnitudes (see **Section 7.4** for more information on this). This suggests that it may be possible to improve outdoor classification performance in high ambient light conditions by modulating training datasets with these types of noise.

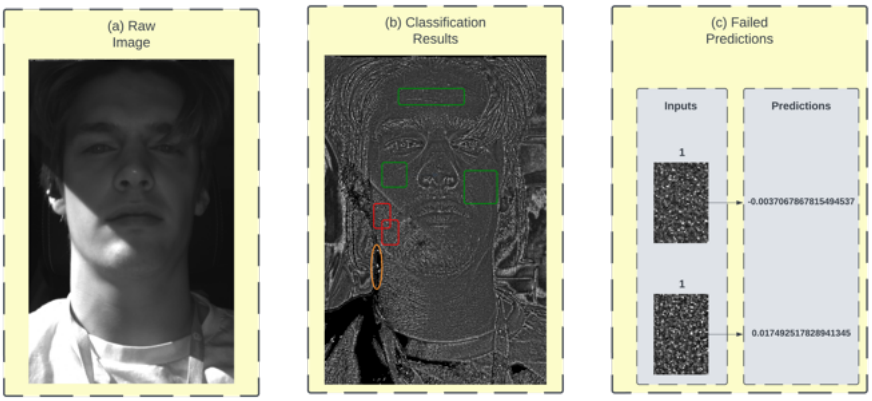

Figure 55: *Example of classification performance for bare skin regions in the 'sunny' ambient light category.* (a) *The source image.* (b) *The LSCI image with successful bare skin classifications shown in green and failed bare skin classifications shown in red. Notably, classification is still effective for regions that are not directly illuminated by sunlight. Overexposed regions of skin are circled in orange.* (c) *The predictions returned for the failed classifications.*

Additionally, **Figure 55** suggests that it may be possible to improve direct sunlight classification performance with lower exposure settings. However, this result also highlights another weakness of this classification technique; given a scene with discontinuous changes in illumination profile, it is not possible to achieve accurate classification results for the entire scene. Instead, a fraction of the scene must be chosen and exposure settings can only be optimised for that area of the scene. For example, in **Figure 55**, exposure is optimised for the shadowed portion of the scene. This is at the trade-off of overexposure on the left side of the chin and neck. While exposure could be adjusted to optimise classification for these regions, doing so would result in the rest of the face being underexposed. Notably, from another perspective this may be seen as a good result as overexposed regions would be undesirable for spoofing detection of biometric signal monitoring.

When interpreting the results of the outdoors data capture, it is important to consider that the classification model was not trained on any data recorded outside. Although light space augmentation was used in an effort to mimic outdoor data, all of the training data was recorded indoors in a controlled laboratory environment with no ambient light.

Finally, it is important to note that the outdoor evaluation dataset was fairly limited, with just 1953, 351, 120 and 461 images for the heavily overcast, overcast, cloudy sunny and sunny categories, respectively. Additionally, a single subject was used for the outdoor evaluation datasets. Despite these limitations, the results still provide valuable insight into the susceptibility of the model to ambient light.

7.4 Performance versus Image Noise

In this section, the model is evaluated against test datasets modulated by different types of noise. This work is conducted in order to gain insight into the effect of image noise on classification performance. For each type of noise that is investigated, examples are provided of where such interference may be encountered in the real world.

7.4.1 Gaussian Noise

Gaussian noise is present within all digital images. Sources of Gaussian noise include poor illumination, high temperature and electronic circuit noise. In this investigation, the test dataset was modulated with Gaussian noise of varying amplitude. This was accomplished by creating a multiplicative masking array for each image in the test dataset. Values in the masking array were randomly sampled from a Gaussian distribution with a mean of one and a standard deviation equal to the desired magnitude of noise modulation. Accordingly, the histogram for each of these masking arrays matched the shape of the specified Gaussian distribution (e.g. **Figure 56**). An example input and it's corresponding Gaussian noised outputs are presented in **Figure 57**.

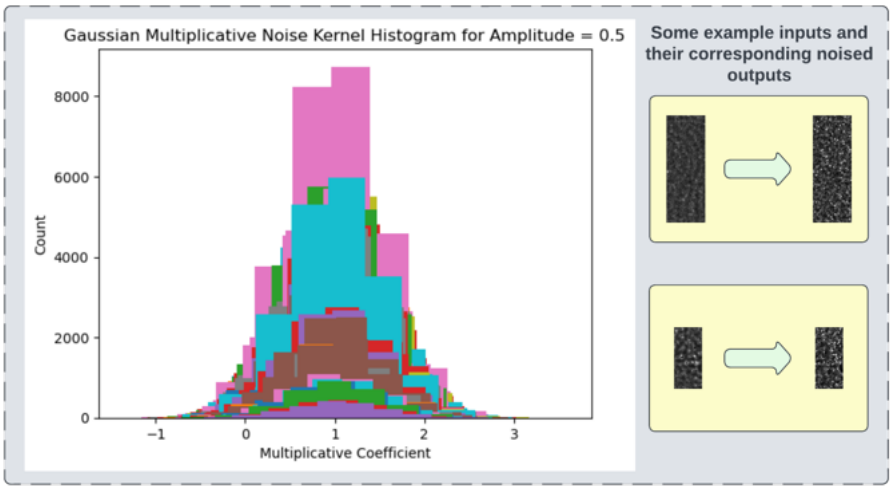

Figure 56: *(left) Histograms of each multiplicative Gaussian noise kernel applied to the test dataset for amplitude equal to 0.5. (right) Some example inputs modulated by a multiplicative Gaussian noise kernel sampled from the distribution described on the left.*

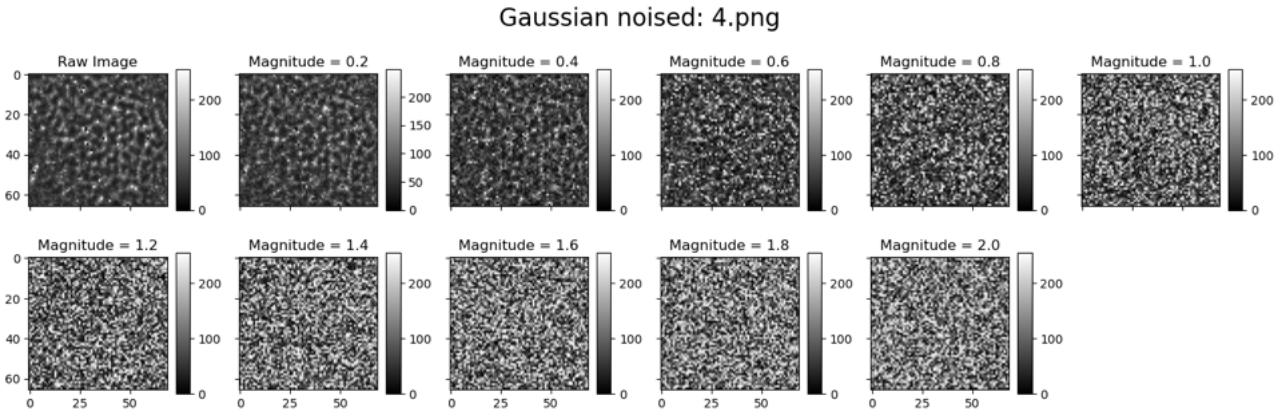

Figure 57: *An example input noised with Gaussian noise of varying magnitude.*

The results of this investigation are summarised in **Figure 58**.

Confusion matrix statistics can be used to provide further insight into the degradation in performance observed as Gaussian noise amplitude increases (**Table 5**).

The confusion matrix statistics show that the models bias towards negative predictions (i.e. 'not skin') increases as Gaussian noise magnitude is increased.

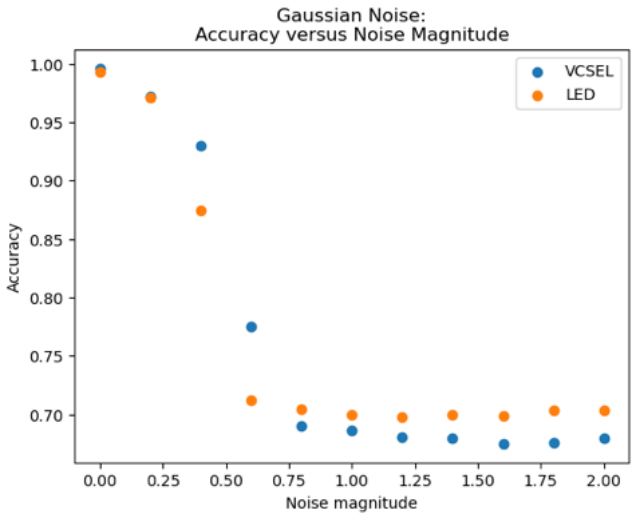

Figure 58: *Accuracy versus Gaussian noise modulation amplitude for the VCSEL and LED models.*

Table 5: *VCSEL model confusion matrix statistics for test dataset modulated by Gaussian noise.*

Magnitude	True Positive	True Negative	False Positive	False Negative
0.0	307	667	4	0
0.2	303	648	23	4
0.4	290	619	52	17
0.6	106	652	19	201
0.8	5	670	1	302
1.0	1	670	1	306
1.2	0	666	5	307
1.4	1	664	7	306
1.6	0	660	11	307
1.8	1	660	11	306
2.0	1	664	7	306

7.4.2 Salt-and-pepper Noise

Salt-and-pepper noise is sometimes found within digital images due to hot pixels or errors in data transmission. Salt-and-pepper noise appears as sparsely occurring black and white pixels. In this investigation, for each test dataset image, a proportion of pixels (equal to the desired amplitude) were randomly selected and written to white (255) or black (0). For example, if an amplitude of 0.1 was passed, 10% of pixels within each test dataset image are affected. An example input and it's corresponding salt-and-pepper noised outputs are presented in **Figure 59**.

The results of this investigation are summarised in **Figure 60**

Similar to **Section 7.4.1**, confusion matrix statistics can be used to provide further insight into the degradation in performance observed as salt-and-pepper noise amplitude increases (**Table 6**).

The confusion matrix statistics are similar to those seen in **Section 7.4.1**, demonstrating a bias towards negative predictions (i.e. 'not skin') as salt-and-pepper noise is increased.

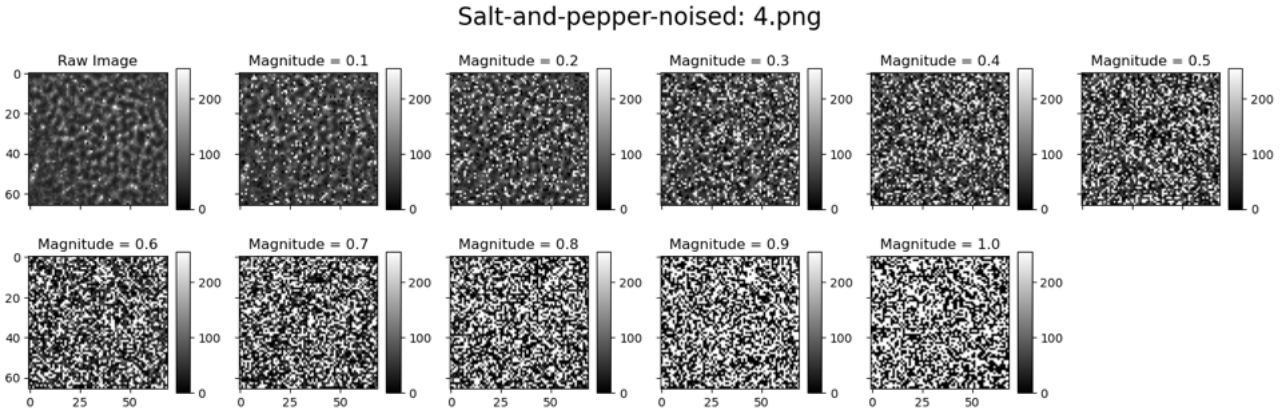

Figure 59: *An example input noised with salt-and-pepper noise of varying magnitude. In the case of magnitude = 1.0, all signal present in the raw image is lost and the resultant image is pure noise.*

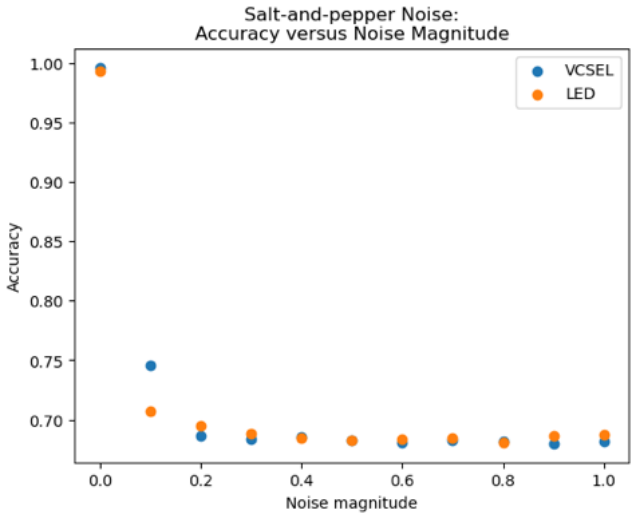

Figure 60: *Accuracy versus salt-and-pepper noise modulation amplitude for the VCSEL and LED models.*

7.4.3 Poisson Noise

Poisson noise, also known as photon shot noise, is present within all digital images. Poisson noise stems from the irregular interval of arrival of photons at the image sensor plane. Poisson noise is most prevalent in low-light images that are multiplied by a digital gain factor. Comparatively, for large amplitudes, Poisson noise is indistinguishable from Gaussian noise. In this investigation, the test dataset was modulated with Photon noise of varying amplitude. This was accomplished in the same manner as described in **Section 7.4.1**, except multiplicative coefficients in the masking array were instead sampled from a Poisson distribution with the desired amplitude of noise modulation used to describe the expectation value of the Poisson distribution. Accordingly, the histogram for each of these masking arrays matched the shape of the specified Poisson distribution (e.g. **Figure 61**). An example input and it's corresponding Poisson noised outputs are presented in **Figure 62**.

The results of this investigation are summarised in **Figure 63**

Similar to **Section 7.4.1**, confusion matrix statistics can be used to provide further insight into the degradation in performance observed as Poisson noise amplitude increases (**Table 7**).

Table 6: *VCSEL model confusion matrix statistics for test dataset modulated by salt-and-pepper noise.*

Magnitude	True Positive	True Negative	False Positive	False Negative
0.0	307	667	4	0
0.1	54	666	5	253
0.2	1	670	1	306
0.3	0	670	1	307
0.4	0	668	3	307
0.5	0	668	3	307
0.6	0	667	4	307
0.7	0	668	3	307
0.8	0	666	5	307
0.9	0	660	2	307
1.0	0	668	3	307

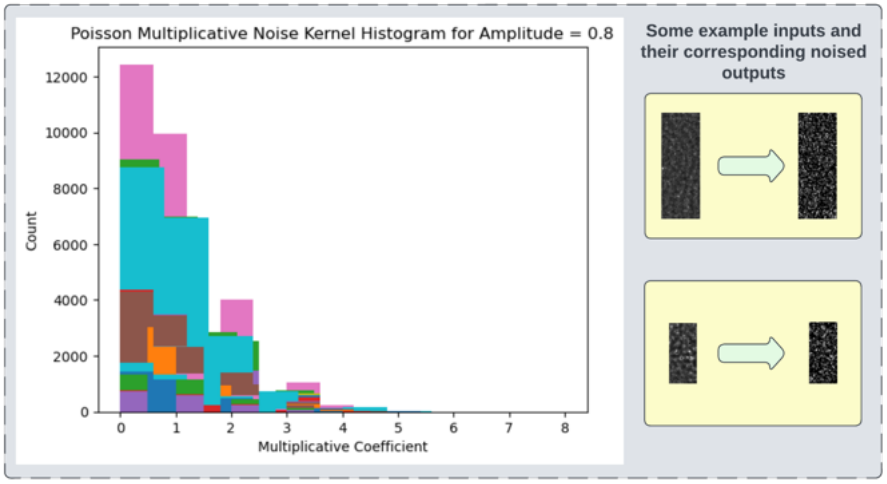

Figure 61: *(left) Histograms of each multiplicative Poisson noise kernel applied to the test dataset for amplitude equal to 0.8. (right) Some example inputs modulated by a multiplicative Poisson noise kernel sampled from the distribution described on the left.*

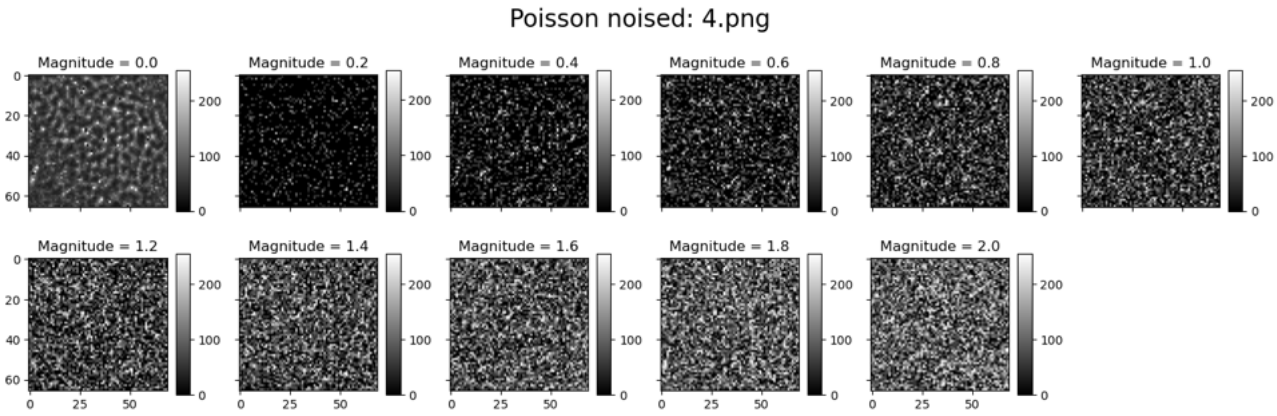

Figure 62: *An example input noised with Poisson noise of varying magnitude.*

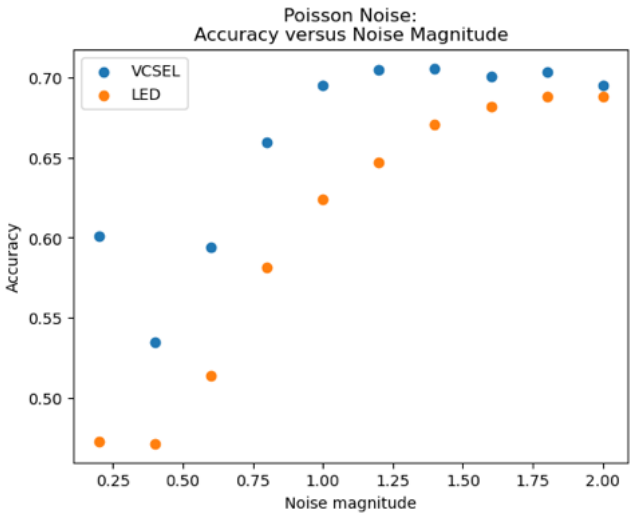

Figure 63: *Accuracy versus Poisson noise modulation amplitude for the VCSEL and LED models.*

Table 7: *VCSEL model confusion matrix statistics for test dataset modulated by Poisson noise.*

Magnitude	True Positive	True Negative	False Positive	False Negative
0.2	73	526	145	234
0.4	159	349	322	148
0.6	175	401	270	132
0.8	161	475	196	146
1.0	138	542	129	169
1.2	117	580	91	190
1.4	100	589	82	207
1.6	88	599	72	219
1.8	73	606	65	234
2.0	55	629	42	252

The confusion matrix statistics are similar to those seen in **Section 7.4.1**, demonstrating a bias towards negative predictions (i.e. 'not skin') as Poisson noise is increased.

7.4.4 Noise Evaluation Conclusion

Two notable findings emerge from the noise evaluation investigation. Firstly, predictions output by the models consistently saturate towards 'not bare skin' as noise magnitude increases. This indicates that the models associates high-frequency information with the 'not bare skin' category. Secondly predictions saturate towards 'bare skin' for low magnitudes of Poisson noise. This suggests that the models associates low-frequency information with the 'bare skin' category.

8 Conclusion

The project was framed around the question: *can the interference patterns produced by coherent light be used for skin detection within the constraints defined by a DMS application?*

In summary, an investigation was completed into the intrinsic properties of coherent light and how these properties may be useful for skin detection in a DMS application context. A comprehensive literature review was conducted examining existing research on the topic of skin detection that could be beneficial in addressing the projects core problem statement. Three potentially viable methods of skin detection were identified: beam profile analysis, laser speckle variation analysis, and laser speckle contrast imaging. An investigation was conducted into the optical properties of skin and a viable coherent light source (VCSELs) was identified that could be easily integrated into the optical pathway of an existing DMS system. Using this modified version of a traditional DMS optical pathway, attempts were made to implement each of the three potentially feasible skin detection methods that were identified in the literature review.

The first of these methods, beam profile analysis, produced very promising initial results. Effective clustering was performed isolating skin from an MDF background. This clustering was done using a simple binary thresholding operation of the standard deviation of each beam profile. Results for beam profile analysis were less compelling when presented with a scene containing a wider range of materials. Ultimately, this investigation was closed due to a lack of available hardware.

The second of these methods, laser speckle variation analysis, was attempted while keeping within the constraints defined by a DMS applications context. In order to obey these constraints, a VCSEL producing a relatively small (in comparison to traditional laser speckle variation analysis implementations) amount of laser speckle was used and image processing techniques were applied in an effort to isolate the speckle from the source images. The utilised image processing techniques were effective in isolating laser speckle, however, they also output a number of other signals present in the source image: namely specular reflection. Ultimately, this meant that attempts to implement laser speckle variation analysis for clustering of skin were unsuccessful.

Finally, LSCI was used in an attempt to produce vein maps. Keeping with the constraints of the project application context, a VCSEL producing a small amount of speckle was again used and the same aforementioned image processing techniques were applied to isolate the speckle (and specular reflection) present in source images. A temporal LSCI computation was then computed over a range of these images. The result of this filtering operation was not successful in producing vascular maps (as initially intended), however, it was effective at increasing the contrast between different textures present within the source images. Notably, visually indistinguishable increases in textural contrast were observed for LED and VCSEL input source images, suggesting that specular reflection was the primary source of signal for the LSCI computation. Despite this, the background theory justifying the reasoning for attempting LSCI LSI processing hypothesised that LSCI images computed using VCSEL source images should provide a consistent increase in classification performance due to the additional classification signal provided by laser speckle.

Careful thought was put into designing a data collection process that would allow for a meaningful comparison to be made between VCSEL and LED illuminated source images such that this hypothesis may be tested. Data collection was performed for 23 different subjects. Datasets were compiled for temporal resolutions 5 through 60 (inclusive) in increments of 5, and for both VCSEL and LED illuminated scenes. For each dataset, a total of 1400 LSCI images were created. From these LSCI images, 9,791 unique ROIs were individually masked by hand.

Supervised learning was then used to create an algorithm capable of performing classification

of regions masked from these LSCI images. More specifically, transfer-learning was used to fine-tune ConvNeXt CNNs on these datasets. Initial results using the test datasets demonstrated near perfect classification scores. Additionally, the results verified the experimental hypothesis, with a consistent increase in classification performance observed for the VCSEL LSCI classifiers. PCA was used to visualise this difference, further reinforcing this finding.

These results did not provide any insight into the in-field viability of the model due to the many assumptions baked into the test datasets. In order to evaluate the impact of these assumptions, a performance evaluation was conducted examining the impact of temporal resolution, image resolution, and ambient light on classification performance. Quantitative results were presented on the impact of each of these assumptions. These results reinforced the aforementioned hypothesis, with VCSEL LSCI classifiers continually outperforming LED LSCI classifiers. The findings of these investigations concluded that: (1) there is no benefit using temporal resolution greater than 25, (2) VCSEL classifiers are significantly more robust to decrease in image resolution than LED classifiers, (3) a more diverse set of training data could be used to greatly improve classification performance for lower image resolutions, (4) classification accuracy of the VCSEL model is not impacted by low ambient light conditions, (5) classification accuracy of the VCSEL model is moderately impacted by conditions with medium ambient light, and (6) predictions output by the VCSEL model become biased towards 'not bare skin' for regions illuminated by direct sunlight.

Finally, an investigation was performed into the impact of various types of image noise on classification performance. The results of this investigation suggested that test datasets modulated by moderate amounts of Gaussian, Poisson and salt-and-pepper noise could be used to improve outdoor classification performance. They also suggested that 'not bare skin' predictions are associated with large amounts of high-frequency image content and 'bare skin' predictions are associated with low-frequency image content.

Summarising these results, **the project demonstrated that the interference patterns produced by coherent light provide a consistent boost in skin detection binary classification performance when implementing LSCI LSI image processing in a DMS application context**. The project was inconclusive in determining whether classification is possible in a DMS system using only these interference patterns.

On a personal level, this project marks a significant increase in skillset. Some of these new skills include:

- Application of computer vision image processing techniques in Python including convolutional filtering operations, frequency domain filtering operations and image resolution manipulation techniques.
- Practice creating large datasets while identifying and maintaining control variables throughout the entire data collection process.
- Practice managing large datasets. In particular, lots of insight was gained into the importance of long-term thinking when organising collected data.
- Development of a rudimentary understanding of photonics and optics. This knowledge was critical in designing an experimental procedure that allowed for a meaningful comparison to be made between results obtained from VCSEL versus LED illuminated source images.
- Practice working in an optical laboratory environment using optics equipment to design custom optical pathways.
- Ability to perform surface mount soldering with the aid of a microscope.

- Experience gained working in a research-oriented role.
- Machine-learning design skills. More specifically, implementation of transfer learning in PyTorch to train a binary regression classification network.
- Machine-learning evaluation skills. More specifically, use of confusion matrix statistics, ROC analysis, precision recall analysis, ablation analysis, and augmentation percentage derivation analysis to evaluate a binary regression classifier.
- Machine-learning insight skills. Use of model kernels, PCA, noising of inputs and comparison of carefully designed test datasets to ascertain insight into what a machine-learning algorithm has actually learnt.

In particular, I found "working in a research-oriented role" to be the most difficult of these skills to develop. This was largely attributable to the hyper-specific nature of research-based work which makes it difficult to find resources and people knowledgeable on the topic of work. Contrastingly, for all other skills listed above it was easy to find existing documentation or knowledgeable experts who I could approach for assistance.

Existing skills that have seen substantial growth and improvement include:

- Development of good coding practices that result in robust and flexible code.
- Development of experience gained working in an professional engineering workplace.
- Effective scoping, design and conduct of a long-term research project.

9 Future Works

Although the findings of the project were in support of the projects core problem statement, a number of questions remain unanswered.

Firstly, a more comprehensive investigation is needed determining the impact of ambient light on the classification performance of the produced models. Rather than using just four qualitative categories for ambient light, this investigation should plot the classification performance of the model as a direct function of ambient light. Such an investigation may be performed by synchronising readings of ambient light with data collection recordings and plotting the classification performance of a given frame as a direct function of ambient light intensity. Careful consideration will need to be given as to how ambient light should be measured, such that consistent recorded values of ambient light will be returned regardless of the position of the sun. Additionally, work should be undertaken to investigate the impact of ambient light on the performance of the LED classifiers.

Secondly, circular polarisers could be used to eliminate specular reflection within the output of the noise isolation function. This would allow for the proportion of specular reflection within the output of the noise isolation function to be quantified. Depending on how this impacts the output of the LSCI filtering operation, a dataset could then be compiled to train a new classification algorithm. Theoretically, this should provide an increase in the robustness of the VCSEL classification algorithm. Additionally, it should result in a significant discrepancy between the classification performance of the VCSEL and LED classifiers. Careful consideration will need to be given as to how irradiance would be matched between scenes illuminated by circularly polarised coherent (VCSEL) and incoherent (LED) light such that a meaningful comparison can be made between classification results.

Thirdly, the findings of the image resolution evaluation (**Section 7.2**) indicate that a more comprehensive training dataset that includes recording taken at different distances could be used to improve the classification performance of the model. The downside of this is that such a dataset would be considerably more tedious to label as an expedited masking process as described in **Section 6.3** would no longer be viable.

Fourthly, the dataset compiled in **Section 6.1** contained only a single female subject. Accordingly, little insight is provided by the results on the viability of the model when used on female subjects. Thus, it would be beneficial for future works to examine how this affects classification performance.

Fifthly, further investigation of the spatial LSCI technique would be beneficial due to the associated low processing requirements and resilience to motion blur. Implementation of this classification method using spatial LSCI (rather than temporal) would significantly increase the attractiveness of this method in to DMS/OMS companies. It is likely a more speckly VCSEL is required to implement this technique.

Sixthly, work is needed to optimise the created classification model such that it is able to run in real time on an embedded system. Additionally, it would be beneficial to implement the algorithm with the existing facetracker defined **Section 1** such that masking of ROIs for classification is performed automatically.

Lastly, initial results obtained using the beam profile analysis method (**Section 5.1**) were very promising. These results justify further investigation of this technique with more capable hardware.

Acknowledgements

First and foremost, I would like to thank the Optics Department of Seeing Machines for their sponsorship of this project. Specific mentions go to Lachlan Whichello for his role as the conceiver of the project's core problem statement as well as his many hours spent advising me on the project, Charles Crespin for his continued support and encouragement, Pratyush Sahay for his critical guidance on the machine-learning side of this project, and Tristan O'Brien who was essential in virtually every hardware component of the project.

10 References

References

Tefft, B.C. (2014) *Prevalence of Motor Vehicle Crashes Involving Drowsy Drivers, United States, 2009-2013*. [Online]. Available at: aaafoundation.org [Accessed 1 Nov. 2022].

European Commission. (2022) *New rules to improve road safety and enable fully driverless vehicles in the EU*. [Online]. Available at: ec.europa.eu [Accessed 1 Nov. 2022].

Australasian New Car Assessment Program. (2022) *Three new EV models bring five star safety*. [Online]. Available at: www.ancap.com.au [Accessed 1 Nov. 2022].

Valuates Reports. (2022) *Global and United States Automotive Driver Monitoring System (DMS) Market Report Forecast 2022-2028*. [Online]. Available at: reports.valuates.com [Accessed 1 Nov. 2022].

Feynmann, R. *The Strange Theory of Light and Matter*, 1985. pp. 85, 100, 113-114 [Online]. Available at: www.mysearch.org.uk [Accessed 16 Sep. 2022].

Azari, B. (2018) *Bidirectional Texture Functions: Acquisition, Rendering and Quality Evaluation*. University of Weimar. [Image]. Available at: www.researchgate.net [Accessed 22 Sep. 2022].

Stannered. (2007) *Michelson Interferometer Diagram*. [Image]. Available at: commons.wikimedia.org [Accessed 1 May 2022].

Hariharan, P. *Holographic Interferometry*, 1992. pp. 117-128 [Online]. Available at: www.sciencedirect.com [Accessed 11 Nov. 2022].

Micheels, J. *et al.* (1984) *Laser doppler flowmetry. A new non-invasive measurement of microcirculation in intensive care?*. Resuscitation. [Online] Available at: pubmed.ncbi.nlm.nih.gov [Accessed 13 Nov. 2022].

Bell, R.J. *Introductory Fourier Transform Spectroscopy*, 1972. [Online]. Available at: https://www.sciencedirect.com/book/9780120851508/introductory-fourier-transform-spectroscopy [Accessed 15 Nov. 2022].

Weisstein, E.W. (2007) *Fourier transform spectroscopy*. [Image]. Available at: https://scienceworld.wolfram.com/physics/FourierTransformSpectrometer.html [Accessed 14 Nov. 2022].

Bates, J.B. (2002) *Fourier transform spectroscopy*. Oak Ridge National Laboratory. [Online]. Available at: www.sciencedirect.com [Accessed 14 Nov. 2022].

Lennartz C. *et al.*. (2020) *Beam Profile Analysis for 3D imaging and material detection whitepaper*. trinamiX [Online]. Available at: trinamixsensing.com/media [Accessed 1 Dec. 2022].

BASF. (2021) *trinamiX 3D Imaging Solution for Driver Monitoring*. [Online]. Available at: youtube.com [Accessed 1 Dec. 2022].

Hustak, L. (2021) *Absorption and Emission spectra*. [Image]. Available at: webbtelescope.org [Accessed 1 May 2023].

Kilgore, G.A., Whillock, P.R. (2008) *Skin Detection Sensor Patent*. Honeywell. [Online]. Available at: patents.google.com [Accessed 14 Nov. 2022].

Li D.Y. *et al.*. (2021) *Transmissive-detected laser speckle contrast imaging for blood flow monitoring in thick tissue: from Monte Carlo simulation to experimental demonstration*, Light Sci Appl. [Online]. Available at: www.nature.com [Accessed 2 Jan. 2023].

Dolan, J.P. *et al.*. (2020) *Qualitative comparison of speckle image processing techniques for vein detection in plant leaf tissue*. SPIE. [Online]. Available at: spiedigitallibrary.org [Accessed 3 Jan. 2023].

Engelen, R. Stasser, E. (2018) *Patent: Laser speckle analysis for biometric authentication*. Available at: patentguru.com [Accessed 1 Sep. 2023].

Jensen, H.W. *et al.*. (2001) *A practical model for subsurface light transport*. Stanford. [Online]. Available at: graphics.standford.edu [Accessed 3 May. 2023].

Albiol, A. *et al.*. (2001) *Optimum color spaces for skin detection*. IEEE. [Online]. Available at: www.researchgate.net [Accessed 10 May 2023].

Jones, M.J. *et al..Statistical Color Models with Application to Skin Detection*, 2002. International Journal of Computer Vision. pp.81-96

Mahoodmi, M.R. *et al.*. (2016) *A Comprehensive Survey on Human Skin Detection*. International Journal of Image, Graphics and Signal Processing. pp. 1-35 [Online]. Available at: mecs-press.org [Accessed 20 Oct. 2022].

Vijayalakshmi M.M. (2019) *Melanoma Skin Cancer Detection using Image Processing and Machine Learning*. ITJTSRD. [Online]. Available at: cloudfront.net [Accessed 20 May 2023].

Kumar, V.B. *et al.*. (2016) *Dermatological disease detection using image processing and machine learning*. Third Internation Conference on Artificial Intelligence and Pattern Recognition. Available at: ieeexplore.ieee.org [Accessed 20 May 2023].

Gajinov, Z. *et al.*. (2010) *Optical properties of the human skin*. Serbian Journal of Dermatology and Venereology. [Online]. Available at: sciendo.com [Accessed 23 Mar. 2022].

Angelopoulou, E. (1999) *The Reflectance Spectrum of Human Skin*. University of Pennsylvania. [Online]. Available at: respository.upenn.edu [Accessed 16 Mar. 2022].

Jacques, S.L. (2013) *Optical properties of biological tissues: a review*. IOP Publishing. [Online]. Available at: omlc.org [Accessed 20 Mar. 2022].

Konom T., Yamada, J. (2019) *In Vivo Measurement of Optical Properties of Human Skin for 450–800 nm and 950–1600 nm Wavelengths*. International Journal of Thermophysics. [Image]. Available at: link.springer.com [Accessed 20 Mar. 2023].

ENLITECH. (2020) *Characterisation on DBR structure of VCSEL laser cavity by u-PLM-EX*. [Image]. Available at: www.enlitechnology.com [Accessed 25 May 2020].

Mandre, S.K. *et al.*. (2008) *Evolution from modal to spatially incoherent emission of a broad-area VCSEL* . Optics Express. [Online]. Available at: opg.optica.org [Accessed 15 May. 2023].

Chemix. (2023) *Online Experimental Diagram Maker*. [Online]. Available at: chemix.org [Accessed 27 May 2023].

Plested, J. Gedeon, T. (2022) *Deep transfer learning for image classification: a survey*. UNSW. [Online]. Available at: arxiv.org [Accessed 20 Feb 2023].

Zhuang Liu, *et al.*. (2022) *A ConvNet for the 2020s*. Facebook AI Research. [Online]. Available at: arxiv.org [Accessed 29 Jan 2023].

Dosovitskiy, A. *et al.*. (2021) *AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE*. Google Research, Brain Team. [Online]. Available at: arxiv.org [Accessed 29 Jan 2023].

Deininger, L. *et al.*. (2022) *A comparative study between vision transformers and CNNs in digital pathology*. Hoffmann-La Roche AG. [Online]. Available at: arxiv.org [Accessed 29 Jan 2023].

Jain, A.K., Farrokhnia, F. (1991) *Unsupervised texture segmentation using Gabor filters*. Michigan State University. [Online]. Available at: www.sciencedirect.com [Accessed 4 Mar. 2022].

Jain, A.K., Dubes, R.C. *Algorithms for clustering data*, 1988. Michigan State University. [Online]. Available at: homepages.inf.ed.ac.uk [Accessed 12 Mar. 2022].

A Fuji and Generalised Differences LSI Image Processing

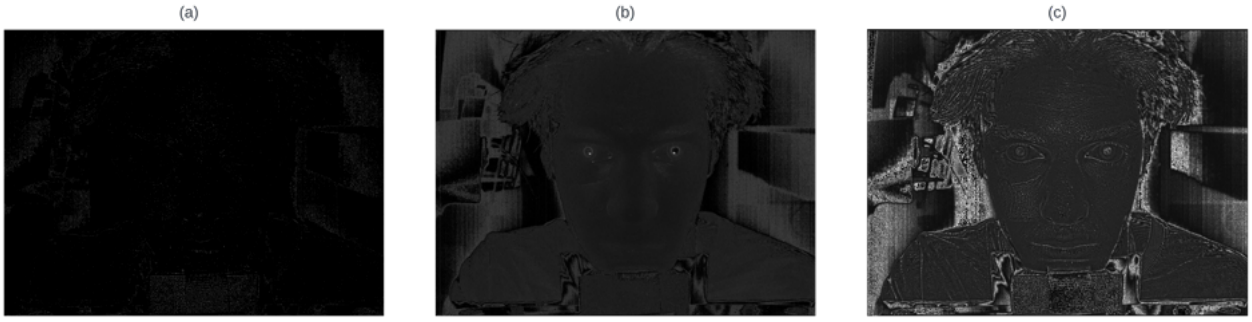

Figure 64: (a) *Fuji computation using 60 source images processed by the noise isolation function as inputs.* (b) *Fuju computation with gain=7 using 60 source images with no pre-processing as inputs.* (c) *Corresponding LSCI computation using same source images.*

Figure 65: (a) *Generalised Differences computation using 60 source images processed by the noise isolation function as inputs.* (b) *Generalised Differences computation using 60 source images with no pre-processing as inputs.* (c) *Corresponding LSCI computation using same source images.*

B Unsupervised Texture Segmentation

Texture analysis is a relatively stale field with minimal progress over the last two decades. This is largely because the diversity of natural and artificial textures makes it impossible to give a universal definition to texture. Traditional methods of texture segmentation invoke a multi-channel filtering approach where a bank of Gabor (or equivalent) filters are used to isolate specific bands of frequency content from an image [37]. Clustering is then performed over each filtered image and the clusters identified in each channel are integrated to perform a segmentation of the source image.

To perform unsupervised clustering of the generated LSCI image, a segmentation approach was adapted from Jain and Farrokhnia's 1991 paper, *Unsupervised Texture Segmentation using Gabor Filters* [37]. A total of 16 real-valued, even-symmetric Gabor filters at four different orientations were used to decompose the source image. An example of such a decomposition can be seen in **Figure 64**.

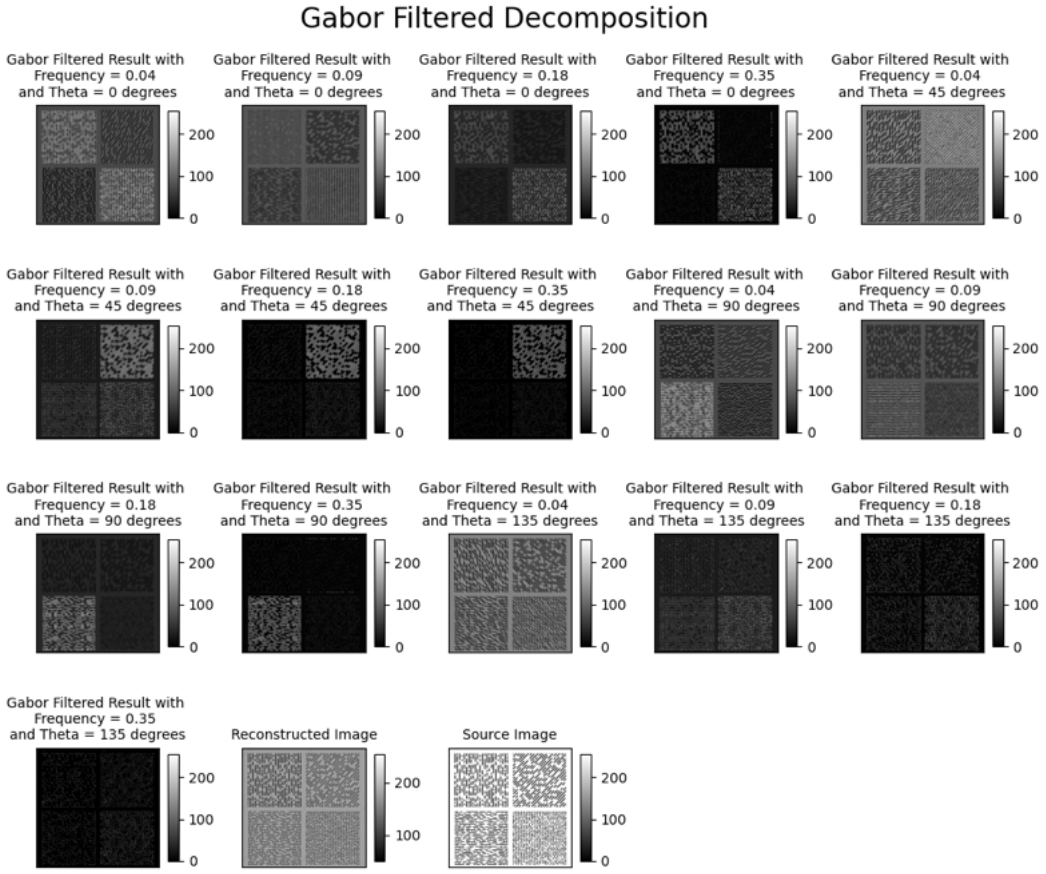

Figure 66: *Gabor filter decomposition of an image consisting of four simple textures, with a bank of 26 filters used.*

The Gabor filter bank does not preserve all information present in the source image. Furthermore, the decomposition is dyadic, meaning lower frequency information is preserved with a higher degree of resolution than high frequency information. The spatial frequency power spectrum of the filter bank can be seen in **Figure 67**. Despite this, majority of features present in the source image are preserved through the filtering operation, as can be seen in a

reconstruction of the source image in **Figure 66**. R^2 value can be used to quantify preservation of information contained in the source image (where an R^2 >0.95 indicates acceptable preservation).

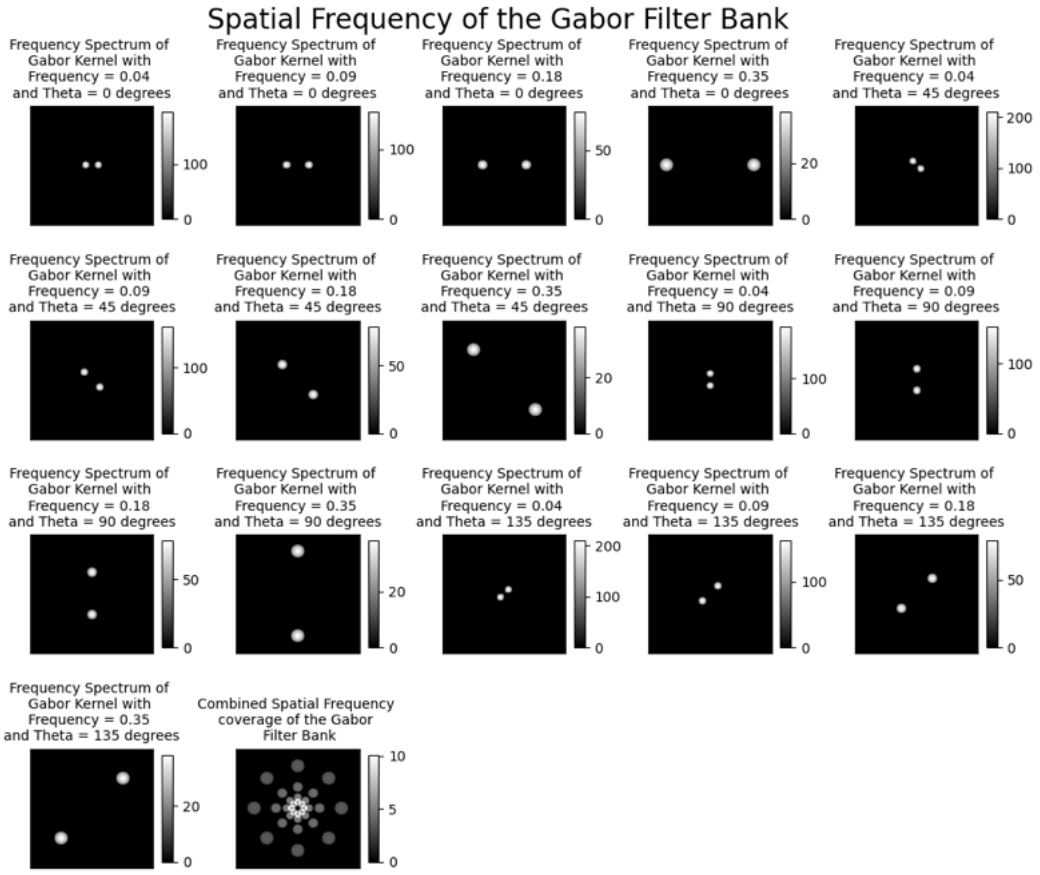

Figure 67: *Spatial frequency coverage of the Gabor filter bank.*

The filtered images are then passed through a non-linear transform with behaviour very similar to that of a sigmoidal activation function. The application of the non-linear transform serves to transform the sinusoidal modulations in the filtered image to square modulations. Finally, in each of the filtered, transformed images, the 'energy' of each distinct texture is computed (**Figure 66**).

Next, clustering is performed on each of the texture energy images so a segmentation may be made. To perform the clustering, a K-means clustering algorithm was created based on Jain and Dubes *Clustering Algorithms and Applications* CLUSTER algorithm [38]. This algorithm consists of a K-means pass followed by a forcing pass.

First, a pattern matrix is formed from an image where pixel intensity and coordinates are stored as features. For the K-means pass, a centroid is initialised at the centre of the pattern matrix and a second centroid is created at the pattern furthest away from the first centroid. Each pattern is then assigned to the centroid with the closest Euclidean distance. The centroids are then recomputed using the new clusterings. A new cluster is then created at the pattern furthest from these two centroids. Each pattern is then reassigned to the closest cluster and the centroids are recomputed. This process is repeated until the desired number of clusters (i.e. clustering) is achieved. This concludes the K-means pass.

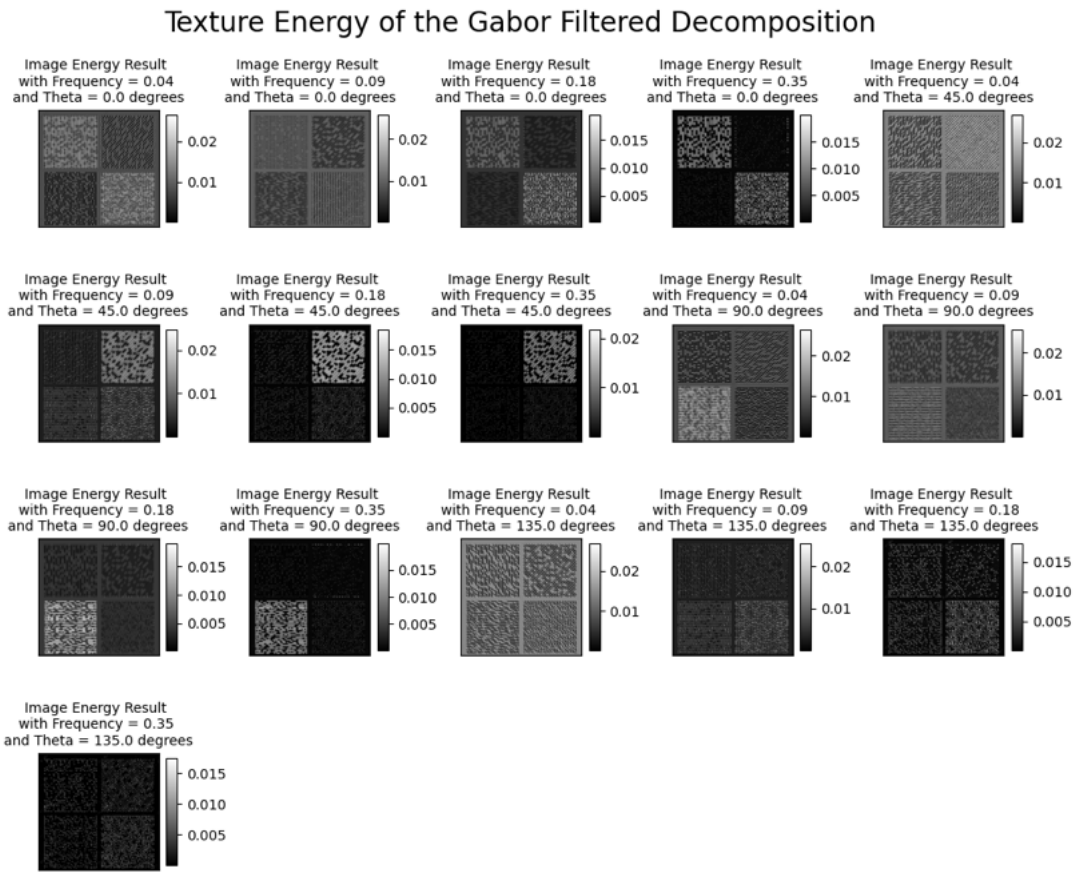

Figure 68: *Textural energy of the Gabor filtered decomposition.*

For the forcing pass, the square error of the first clustering (K=2) is computed. Then, the square error for all possible combinations of the next clustering (i.e. K=3 compressed to K=2) are computed. If any of these square errors are smaller than the square error of the previous clustering, the K-means pass begins again from this new clustering. This process is repeated until the algorithm converges towards the clustering with the smallest square error. An example of a clustering performed by this algorithm can be seen in **Figure 69**.

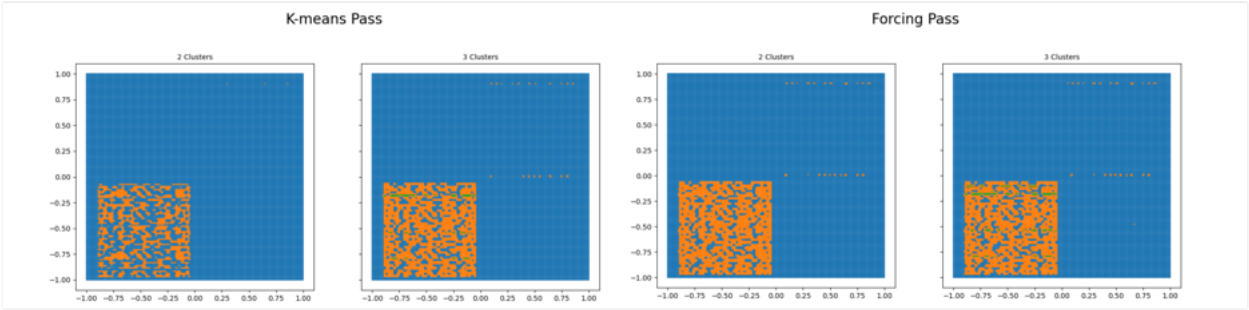

Figure 69: *Resultant clustering of the textural energy image associated with the Gabor filter with frequency = 0.35 and orientation = 90°.*

Figure 69 indicates that the clustering algorithm performs as intended, however, the algo-

rithm is slow to converge, taking anywhere from 4 to 20 minutes for a single image depending on the number of desired clusters. Given that a single image has 16 decompositions when using the filter bank described in **Figure 67**, this would mean the CLUSTER algorithm would need to be run 16 times for each image, before a textural segmentation could be produced. Ultimately, it was deemed that this performance would not be satisfactory for the desired use-case.

It is important to note that, at this stage, no efforts had being made to optimise the performance of the algorithm. Downsampling of source inputs could be used to significantly improve this performance. Additionally, choosing a threshold rather value when comparing the square error between two clusterings (rather than enforcing a hard greater than) could also significantly improve this performance.

Finally, a number of improvement could have been made to the CLUSTER algorithm. In images with many clusterings, pixels at the corners of the image tend to form their own clusters. This is due to the coordinate components of their corresponding patterns dominating the impact of their intensity in the Euclidean distance combination. This could be improved by using the log of each coordinate value. Additionally, a different filter bank could have been selected to provide greater resolution of high-frequency image components.

C ConvNeXt backbone summary

```
ConvNeXt (sequential):
    stem (sequential):
        depthwise convolution: channels_in_out=(3,192), ksize=(4,4), stride=(4,4)
        layer normalisation
    endstem
    stage 0 (sequential):
        3x ConvNeXt block (sequential):
            depthwise convolution: channels_in_out=(192,192), ksize=(7,7), stride=(1,1)
            layer normalisation
            linear: in_features=192, out_features=768, bias=True
            GELU()
            linear: in_features=768, out_features=192, bias=True
        endConvNeXtBlock
    endstage
    stage 1 (sequential):
        downsampling (sequential):
            layer normalisation
            depthwise convolution: channels_in_out=(192,384), ksize=(2,2), stride=(2,2)
        enddownsampling
        3x ConvNeXt block (sequential):
            depthwise convolution: channels_in_out=(384,384), ksize=(7,7), stride=(1,1)
            layer normalisation
            linear: in_features=384, out_features=1536, bias=True
            GELU()
            linear: in_features=1536, out_features=384, bias=True
        endConvNeXtBlock
    endstage
    stage 2 (sequential):
        downsampling (sequential):
            layer normalisation
            depthwise convolution: channels_in_out=(384,768), ksize=(2,2), stride=(2,2)
        enddownsampling
        27x ConvNeXt block (sequential):
            depthwise convolution: channels_in_out=(768,768), ksize=(7,7), stride=(1,1)
            layer normalisation
            linear: in_features=768, out_features=3072, bias=True
            GELU()
            linear: in_features=3072, out_features=768, bias=True
        endConvNeXtBlock
    endstage
    stage 3 (sequential):
        downsampling (sequential):
            layer normalisation
            depthwise convolution: channels_in_out=(768,1536), ksize=(2,2), stride=(2,2)
        enddownsampling
        3x ConvNeXt block (sequential):
            depthwise convolution: channels_in_out=(1536,1536), ksize=(7,7), stride=(1,1)
            layer normalisation
            linear: in_features=1536, out_features=6144, bias=True
            GELU()
            linear: in_features=6144, out_features=1536, bias=True
        endConvNeXtBlock
    endstage
endConvNeXt
```

Algorithm 1: ConvNeXt model architecture internal structure summary

D Non-binary categorical classifier

A non-binary categorial classifier was also trained. For a given input, this model output predictions for each category in the training dataset. The category with the highest prediction score is used as the final classification prediction.

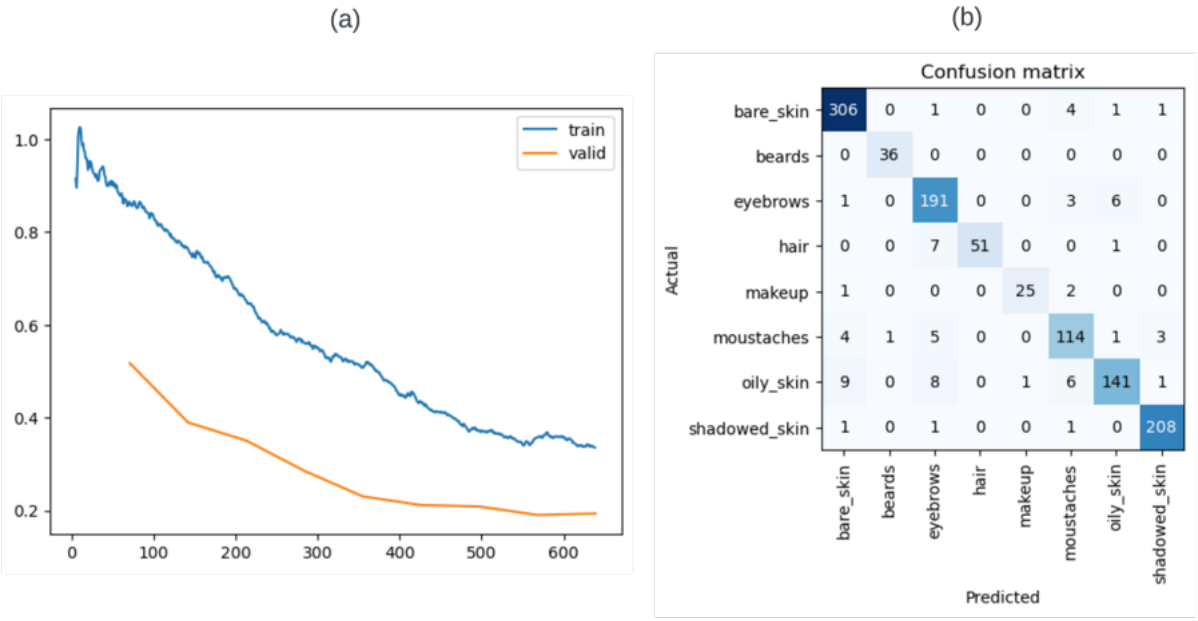

Figure 70: (a) *Training and validation losses for the non-binary categorical classifier trained on source VCSEL LSCI images.* (b) *Classification results for the non-binary categorical classifier on the test dataset.*

E Binary classification with no pre-processing

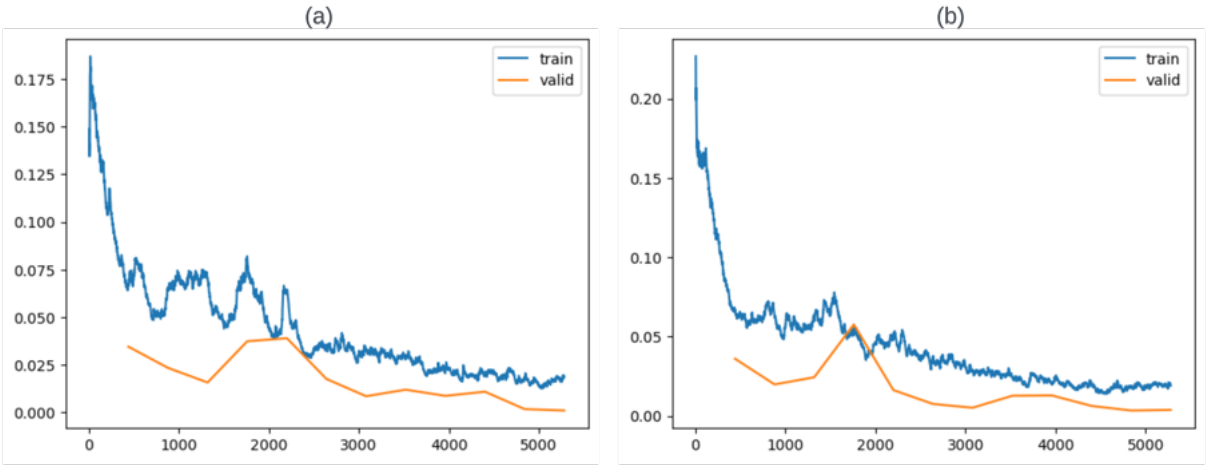

Figure 71: (a) *Training and validation losses for the binary regression LED classifier trained on source LED images with no pre-processing.* (b) *Training and validation losses for the binary regression VCSEL classifier trained on source vCSEL images with no pre-processing.*

In this case, both the LED and VCSEL models achieved a perfect classification score on the test dataset.