Inverse Design Using CNN-LSTM Based Network For Ultra-Low Reflectance Anti-Reflection Coating

A Thesis submitted in partial fulfilment of the requirement for the degree of

Master of Technology

Submitted by

Rahul Prakash

Roll No. 224102314

Supervisor

Dr.Debabrata Sikdar , Dr.Prithwijit Guha

Department of Electronics and Electrical Engineering

Indian Institute of Technology Guwahati

Assam - 781039, India

june 24, 2024

DECLARATION

This is to declare that the work which is being presented in the thesis entitled, **Nanophotonics Inverse Design Using ML and AI Technology** *submitted to the Indian Institute of Technology, Guwahati for the granting of the Master of Technology degree, is a genuine work completed under Dr.Debabrata Sikdar and Dr.Prithwijit Guha.*

I hereby swear that this thesis is entirely original with no submissions made for credit toward another degree or professional certification. I also want to clarify that, to the best of my understanding, there is no plagiarism in this report.

Rahul Prakash
Department of Electronics & Electrical Engineering
IIT Guwahati, India, Assam-781039
june 2024

CERTIFICATE

This is to certify that the work contained in the thesis entitled, **Nanophotonics Inverse Design Using ML and AI Technology**, *is a bonafide work of* **Rahul Prakash (Roll No. 224102314)**, *which has been carried out in the Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati under my supervision and this work has not been submitted elsewhere for a degree.*

Dr.Debabrata Sikdar
Dept. of Electronics and Electrical Engineering
IIT Guwahati, India, Assam-781039
june 2024

ACKNOWLEDGEMENT

I would like to extend my sincere thanks to my supervisor ***Dr.Debabrata Sikdar and Dr.Prithwijit Guha*** *for his constant unwavering support and guidance. Without his support and guidance, it would not have been possible to carry out the project. I also express my gratitude to all the other faculty members of IIT Guwahati's Department of Electronics and Electrical Engineering, whose support and encouragement at important stages gave shape to my project in the present form. Furthermore, I would also like to acknowledge the previous works mentioned in the Bibliography section, which were of tremendous help during the course of the project*

Sincerely

Rahul Prakash

Abstract

In nanophotonics, the integration of artificial intelligence (AI) and machine learning (ML) has unlocked new possibilities for enhancing optical materials and designs. This project focuses on the inverse design of antireflection (AR) coatings, utilizing advanced neural networks to optimize their performance.

Initially, we developed a Python code implementing the Transfer Matrix Method (TMM) to calculate reflectance and transmittance in multilayer structures, specifically targeting AR coatings. Building on this foundation, we created a sophisticated ML model combining Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Multi-Layer Perceptrons (MLP). Our approach addresses the real-world challenge of designing AR coatings for solar cells to minimize reflectance to less than 1 percent The ML model successfully predicts suitable materials and optimal thickness values, demonstrating a significant improvement over conventional methods.

This research highlights the transformative impact of AI on nanophotonics, merging cutting-edge computational techniques with traditional photonics methodologies. The results contribute to academic understanding and pave the way for practical applications in telecommunications, sensors, and imaging technologies.Our findings represent a crucial step in leveraging AI for photonics, showcasing its potential to revolutionize light manipulation and optical performance.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Nanophotonics is a cutting-edge interdisciplinary field that explores the interaction between light and matter at the nanoscale, where dimensions are on the order of a billionth of a meter. This field merges principles from photonics, the study of light, with nanotechnology, the manipulation of materials at the nanoscale. Nanophotonics aims to control and manipulate light on the smallest possible scale, enabling the development of novel devices and applications with unprecedented capabilities. By exploiting the unique properties of materials and structures at the nanoscale, nanophotonics has the potential to revolutionize various areas, including telecommunications, imaging, sensing, and energy harvesting. Researchers in nanophotonics design and engineer structures such as nanoantennas, photonic crystals, and plasmonic devices to achieve enhanced light-matter interactions and create devices that can surpass the limitations of conventional optics. The field's advancements have opened up new possibilities for the development of ultra-compact and high-performance optical components, paving the way for innovations in information processing, medical diagnostics, and beyond. Nanophotonics is at the forefront of pushing the boundaries of what is possible with light, making it a dynamic and exciting area of research with significant implications for the future of technology. [?]

1.1 Basics of Nanophotonics

1.1.1 Fundamentals of Light-Matter Interaction

Light-matter interaction is the fundamental process by which light and matter exchange energy or momentum. This interaction is responsible for a wide range of phenomena, including absorption, emission, transmission, and reflection. In the context of nanophotonics, light-matter interaction is particularly important because it allows us to control light at the nanoscale. This enables a variety of new applications, such as optical communication, sensing, and imaging.

There are two main types of light-matter interaction: classical and quantum. Classical light-matter interaction occurs when light interacts with matter through its electric and magnetic fields. This type of interaction is well understood and is responsible for many of the everyday phenomena that we observe, such as the reflection of light from a mirror or the absorption of light by a black object.

Quantum light-matter interaction occurs when light interacts with matter at the atomic or molecular level. This type of interaction is more complex and is not as well understood. However, it is responsible for some of the most fascinating phenomena in nanophotonics, such as the emission of light by a single molecule or the trapping of light in a nanoscale cavity. Light-matter interaction is a fundamental process in nanophotonics and is responsible for a wide range of phenomena. By understanding and controlling light-matter interaction, we can develop new technologies that will revolutionize the way we interact with light.

1.1.2 Absorption and Emission in Nanophotonics

Nanophotonics is a field of study that focuses on the interaction of light with matter at the nanoscale. This interaction can lead to a variety of phenomena, including absorption and emission.

Absorption is the process by which light is absorbed by matter. When light is absorbed, its energy is transferred to the matter, which can cause it to heat up or undergo a chemical reaction. In nanophotonics, absorption can be enhanced by using nanostructures to trap light in a small volume. This can lead to more efficient light absorption, which is important for applications such as solar cells and photodetectors.

Emission is the process by which light is emitted from matter. When light is emitted, its energy is transferred from the matter to the light field. In nanophotonics, emission can be controlled by using nanostructures to modify the properties of the light field. This can lead to more efficient light emission, which is important for applications such as light-emitting diodes (LEDs) and lasers.

Nanophotonics absorption and emission are fundamental processes in nanophotonics and are responsible for a wide range of phenomena. By understanding and controlling nanophotonics absorption and emission, we can develop new technologies that will revolutionize the way we interact with light.

1.1.3 Reflection and Refraction of Light

Reflection of Light

Once the ray of light strikes a surface of an object, it forms an angle of incidence and the angle of reflection. The Angle of Incidence, The angle which the incident ray forms with the normal line at the point of incidence. The Angle of Reflection, The angle which the reflected ray forms with the normal at the point where the reflected ray bounces back.

Laws of Reflection

(1) The angle of incidence is always equal to the angle of reflection. Both the angle of incidence and the angle of reflection are measured from the normal line $\theta_i = \theta_{\text{reflection}}$.

(2) The incident ray, the normal, and the reflected ray- all lie in the same plane.

(3) The incident ray, as well as the reflected ray, are on the opposite sides of

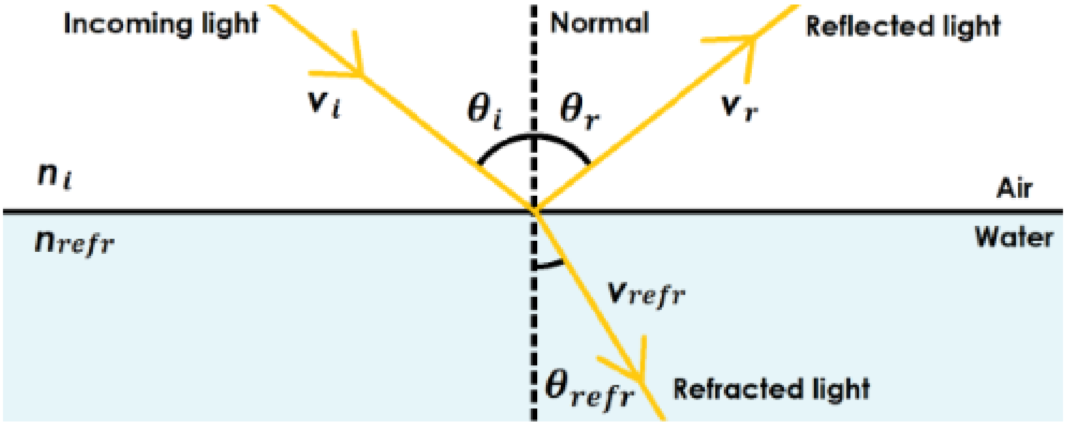

Figure 1.1: Reflection and Refraction

the normal line.

The laws of reflection are applicable to all reflecting surfaces like polished surfaces, plane surfaces (mirrors), and curved surfaces (spherical mirrors).

Refraction of Light

Light travels at the constant speed in space or when the medium is same. The speed of light varies when it travels from one medium to another. When the light enters a glass medium, its speed reduces. When the light strikes the surface of the glass at an angle of 90°, it passes straight through the glass; nonetheless, its speed decreases. However, when the ray of light strikes the surface of a glass at an angle other than 90°, not only does its speed reduces; but the ray of light bends also. This phenomenon of the bending of light, as the light passes from one medium to another, is called 'refraction.'

Laws of Refraction

(1) The incident ray, the normal, and the refracted ray- all lie in the same plane.

(2) Snell's law describes the relationship between the incident ray and the refracted ray. For a pair of two media, the ratio of the sine of the angle of incidence to the sine of the angle of refraction is always constant.

$$\frac{\sin i}{\sin r} = \text{constant}\,(1\mu_2)$$

1.1.4 Polarization of Light

Polarization of light refers to the orientation of the electric field vector associated with an electromagnetic wave. Light waves can vibrate in various directions perpendicular to their direction of propagation. The two most common types of polarization are referred to as "s-polarization" and "p-polarization."
Linear Polarization- In linear polarization, the electric field vector oscillates in a straight line along a specific direction perpendicular to the direction of propagation, Light can be linearly polarized vertically, horizontally, or at any angle in between.
Circular Polarization- Circular polarization occurs when the electric field vector rotates in a circular motion as the wave propagates, Circular polarization can be either right-handed (clockwise rotation) or left-handed (counterclockwise rotation). Elliptical Polarization: Elliptical polarization is a combination of linear and circular polarization, where the electric field vector traces an elliptical path.

S and P Polarization

S-Polarization (Transverse Electric - TE)- In S-polarization, the electric field vector is perpendicular to the plane of incidence, For example, in the context of light reflecting off a surface, if the electric field vector is parallel to the plane formed by the incident ray and the surface normal, it is considered S-polarized, S-polarized light is sometimes called transverse electric (TE) polarization.
P-Polarization (Transverse Magnetic - TM)- In P-polarization, the electric field vector lies in the plane of incidence, For reflection, if the electric field vector is perpendicular to the plane formed by the incident ray and the surface normal, it is considered P-polarized, P-polarized light is sometimes called transverse magnetic (TM) polarization.

1.2 Thin film Reflection

A thin film is a nanometer to micrometer-scale layer of material deposited on a substrate. These films exhibit unique optical, electrical, or mechanical properties. They can be metallic, dielectric, or semiconductor-based. Thin films are commonly used in electronic devices, optics, and coatings. Methods of deposition include physical vapor deposition, chemical vapor deposition, and spin coating. Interference effects in thin films lead to iridescence and anti-reflective coatings. Thin films play a crucial role in technologies like solar cells, displays, and sensors. The thickness and composition of thin films are carefully controlled for specific.

applications. Understanding thin film behavior is essential for tailoring material properties in diverse fields. When light encounters a thin film, it can undergo a variety of phenomena, including reflection, refraction, and interference. The reflection of light from a thin film can be constructive or destructive, depending on the thickness of the film and the wavelength of the light.

Constructive interference occurs when the light waves reflected from the top and bottom surfaces of the film are in phase with each other. This causes the waves to combine and reinforce each other, resulting in a brighter reflection.

Destructive interference occurs when the light waves reflected from the top and bottom surfaces of the film are out of phase with each other. This causes the waves to cancel each other out, resulting in a dimmer reflection. The color of the reflected light depends on the thickness of the film and the wavelength of the light. If the film is thin enough, only certain wavelengths of light will undergo constructive interference, and the reflected light will appear colored. This is why soap bubbles and oil slicks often appear iridescent.

Thin film reflections are used in a variety of applications, including-

Anti-reflective coatings - Anti-reflective coatings are thin films that are designed to reduce the amount of light that reflects from a surface. This is important for applications such as eyeglasses and camera lenses.

Mirrors- Mirrors are thin films that are designed to reflect as much light as

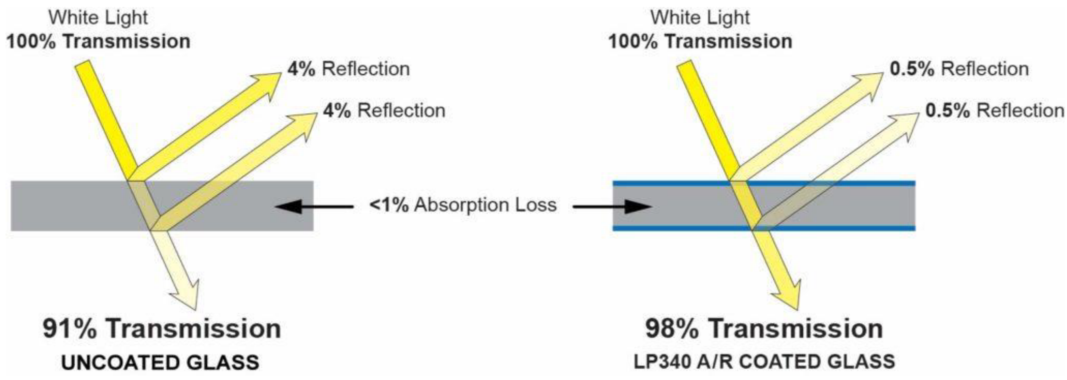

Figure 1.2: AR coating

possible. This is important for applications such as telescopes and microscopes. **Filters-** Filters are thin films that are designed to transmit or block specific wavelengths of light. This is important for applications such as color filters and optical sensors

1.3 Anti-reflective coatings

An anti-reflection coating is a thin film applied to optical surfaces to minimize reflection and enhance transmission of light. Typically composed of dielectric materials, such coatings work by creating destructive interference between reflected waves. This interference reduces glare, ghost images, and increases contrast. Anti-reflection coatings are widely used in eyeglasses, camera lenses, and display screens. The design of these coatings is optimized for specific wavelengths to reduce reflections over a broad spectrum. They improve optical performance by allowing more light to pass through and minimizing unwanted reflections. The effectiveness of anti-reflection coatings is influenced by factors such as film thickness and refractive index [3].

1.3.1 Applications of AR Coating

Eyeglasses and Sunglasses AR coatings on eyeglass lenses reduce reflections, glare, and halos, improving visual clarity and comfort for the wearer.

Camera Lenses and Photographic Equipment AR coatings on camera lenses enhance image quality by minimizing reflections, improving contrast, and preventing ghosting or flare in photographs.

Microscopes and Telescopes AR coatings on optical instruments improve image quality by reducing reflections, allowing for clearer observations and increased contrast in microscopy and astronomy.

Displays (LCD, LED, OLED) AR coatings on screens of electronic devices minimize reflections and improve visibility, providing better viewing experiences for users.

Photovoltaic Panels AR coatings on solar panels enhance light absorption by reducing surface reflections, improving the overall efficiency of the solar cells.

Optical Filters AR coatings are applied to optical filters used in various applications, such as fluorescence microscopy and spectroscopy, to enhance light transmission and minimize reflections.

Medical Devices AR coatings on lenses and optics in medical devices, such as endoscopes and diagnostic equipment, enhance image quality and clarity during medical procedures.

1.4 Multilayers thin film and TMM method

Multilayer thin films are structures composed of multiple layers of different materials, each with specific optical properties, thicknesses, and refractive indices. These films are designed to manipulate the transmission, reflection, and absorption of light for various applications in optics and photonics. The construction and optimization of multilayer thin films involve careful consideration of the properties of each layer and their arrangement.

The transfer-matrix method (TMM) is a computational technique used to analyze the optical properties of multilayer thin films. It involves calculating the propagation of electromagnetic waves through each layer of the film, taking into account the reflection and refraction at each interface. The TMM can be used to predict the reflectance, transmittance, and absorptance of light as a function of wavelength and incident angle [4].

Reflectance is the ratio of the reflected light intensity to the incident light intensity. It is denoted by the symbol R. A reflectance of 0 indicates that no light is reflected, while a reflectance of 1 indicates that all light is reflected. it is calculated using the following formula.

$$R = \left| \frac{r_{21}}{r_{12}} \right|^2 \tag{1.1}$$

where r21 is the amplitude reflection coefficient at the second interface (incident side of the second r12 is the amplitude reflection coefficient at the first interface (side facing the incident medium).

Transmittance is the ratio of the transmitted light intensity to the incident light intensity. It is denoted by the symbol T, A transmittance of 0 indicates that no light is transmitted, while a transmittance of 1 indicates that all light is transmitted. it is calculated using the following formula:

$$T = \left| \frac{t_{21}}{t_{12}} \right|^2 \left(\frac{n_2}{n_1} \right) \tag{1.2}$$

where t21 is the amplitude transmission coefficient at the second interface, t12 is the amplitude transmission coefficient at the first interface, n1 is the refractive index of the incident medium, and n2 is the refractive index of the substrate or the next layer.

Absorptance is the fraction of light that is absorbed by the thin film. It is denoted by the symbol A, The absorptance is related to the reflectance and transmittance by the following equation:

$$A = 1 - R - T \tag{1.3}$$

Transfer Matrix for Each Layer:

The TMM involves calculating the transfer matrix for each layer of the thin film. The transfer matrix for a single layer is a 2×2 matrix that relates the

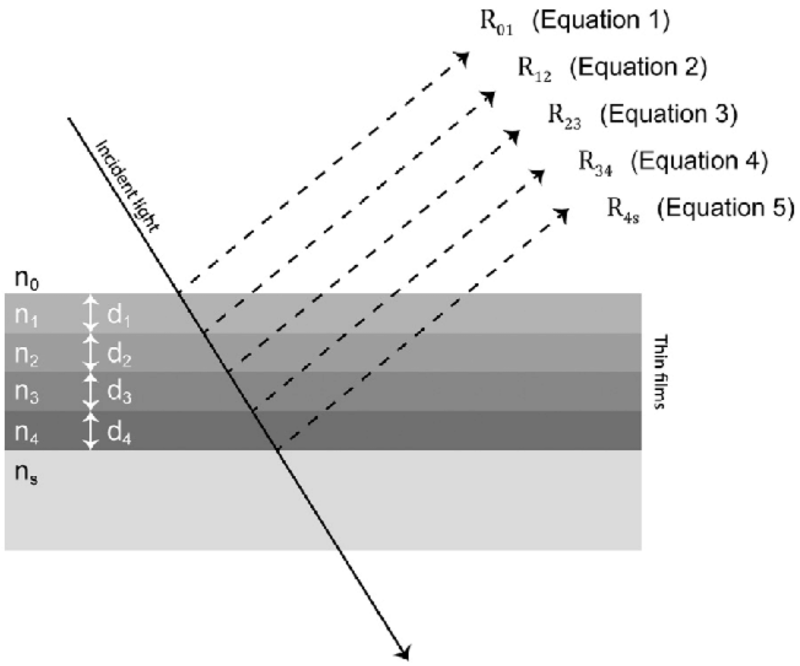

Figure 1.3: Reflection through Multiple layers

electric and magnetic field components of the incident wave to the electric and magnetic field components of the transmitted wave. The transfer matrix for a multilayer thin film is the product of the transfer matrices for each layer.
The TMM involves representing each layer as a 2x2 transfer matrix. For a layer with thickness d, refractive index n, and wave vector k, the transfer matrix M is given by

$$M = \begin{bmatrix} \cos(kd) & \frac{i}{k}\sin(kd) \\ \frac{i}{k}\sin(kd) & \cos(kd) \end{bmatrix} \tag{1.4}$$

Where k is the wave vector, $k = \frac{2\pi n}{\lambda}$ is the wavelength of light in the material. The overall transfer matrix M total for the entire multilayer structure is obtained by multiplying the individual layer matrices

$$M_{\text{total}} = M_1 \cdot M_2 \cdot \ldots \cdot M_{N-1} \cdot M_N \tag{1.5}$$

Where N is the total number of layers.

Chapter 2

AI-ML In Nanophotonics

2.1 What is Artificial intelligence (AI)?

Artificial Intelligence (AI) is a branch of computer science that aims to create intelligent machines capable of mimicking human cognitive functions. It involves the development of algorithms, models, and systems that enable computers to perform tasks that typically require human intelligence. AI encompasses various subfields, including machine learning, natural language processing, computer vision, and robotics. Machine learning, a key component of AI, involves training algorithms on data to improve performance over time. AI applications range from virtual assistants and recommendation systems to autonomous vehicles and medical diagnosis. Deep learning, a subset of machine learning, employs neural networks with multiple layers to enhance the understanding of complex patterns. AI technologies have advanced rapidly, impacting industries such as healthcare, finance, and manufacturing. Ethical considerations surrounding AI include issues of bias, transparency, and accountability. The Turing Test is a classic measure for AI, evaluating a machine's ability to exhibit human-like intelligence. AI is leveraged for problem-solving, decision-making, and automation, transforming how we interact with technology. As AI evolves, ongoing research focuses on achieving artificial general intelligence, enabling machines to perform a wide range of tasks with human-like adaptability.

2.2 Machine Learning And its Type

Machine learning is the branch of Artificial Intelligence that focuses on developing models and algorithms that let computers learn from data and improve from previous experience without being explicitly programmed for every task. In simple words, ML teaches the systems to think and understand like humans by learning from the data. In this article, we will explore about the various types of machine learning algorithms that are important for future requirements. Machine learning is generally a training system to learn from past experiences and improve performance over time. Machine learning helps to predict massive amounts of data. It helps to deliver fast and accurate results to get profitable opportunities.

2.2.1 Types of Machine Learning

There are several types of machine learning, each with special characteristics and applications. Some of the main types of machine learning algorithms are as follows

1. **Supervised Machine Learning**
2. **Unsupervised Machine Learning**
3. **Semi-Supervised Machine Learning**
4. **Reinforcement Learning**

2.2.2 Supervised Machine Learning

Supervised learning is defined as when a model gets trained on a "Labelled Dataset". Labelled datasets have both input and output parameters. In Supervised Learning algorithms learn to map points between inputs and correct outputs. It has both training and validation datasets labelled. Let's understand it with the help of an example. Example: Consider a scenario where you have to build an image classifier to differentiate between cats and dogs. If you feed the datasets of dogs and cats labelled images to the algorithm, the machine will learn to classify between a dog or a cat from these labeled images.

When we input new dog or cat images that it has never seen before, it will use the learned algorithms and predict whether it is a dog or a cat. This is how supervised learning works, and this is particularly an image classification. There are two main categories of supervised learning that are mentioned below

1. Classification

2. Regression

2.2.3 Unsupervised Machine Learning:

Unsupervised Learning Unsupervised learning is a type of machine learning technique in which an algorithm discovers patterns and relationships using unlabeled data. Unlike supervised learning, unsupervised learning doesn't involve providing the algorithm with labeled target outputs. The primary goal of Unsupervised learning is often to discover hidden patterns, similarities, or clusters within the data, which can then be used for various purposes, such as data exploration, visualization, dimensionality reduction, and more.

Example Consider that you have a dataset that contains information about the purchases you made from the shop. Through clustering, the algorithm can group the same purchasing behavior among you and other customers, which reveals potential customers without predefined labels. This type of information can help businesses get target customers as well as identify outliers.

There are two main categories of unsupervised learning that are mentioned below

1. Clustering

2. Dimensionality Reduction

2.2.4 Semi-Supervised Learning

Semi-Supervised learning is a machine learning algorithm that works between the supervised and unsupervised learning so it uses both labelled and unlabelled data. It's particularly useful when obtaining labeled data is costly, time-consuming, or resource-intensive. This approach is useful when the dataset is expensive and time-consuming. Semi-supervised learning is chosen when labeled data requires skills and relevant resources in order to train or learn from it. Let's understand it with the help of an example.

Example Consider that we are building a language translation model, having labeled translations for every sentence pair can be resources intensive. It allows the models to learn from labeled and unlabeled sentence pairs, making them more accurate. This technique has led to significant improvements in the quality of machine translation services.

2.2.5 Reinforcement Machine Learning

Reinforcement machine learning algorithm is a learning method that interacts with the environment by producing actions and discovering errors. Trial, error, and delay are the most relevant characteristics of reinforcement learning. This methods allows machines to automatically determine the ideal behaviour within specific context in order to maximize performance. This type of learning is crucial for applications that involve decision-making in unpredictable environments. Let's understand it with the help of examples.

ExampleConsider that you are training an AI agent to play a game like chess. The agent explores different moves and receives positive or negative feedback based on the outcome. Reinforcement Learning also finds applications in which they learn to perform tasks by interacting with their surroundings [5].

2.3 Neural Networks

A neural network is a computational model inspired by the structure and functions of the human brain. It consists of interconnected artificial neurons that process information, enabling it to learn the patterns and correlations among data points and make predictions from data. Neural networks are used for various tasks, including pattern recognition, decision-making, and solving complex problems in machine learning and artificial intelligence fields. An artificial neural network having several layers of connected artificial neurons is known as a Deep Neural Network (DNN), also sometimes referred to as a deep

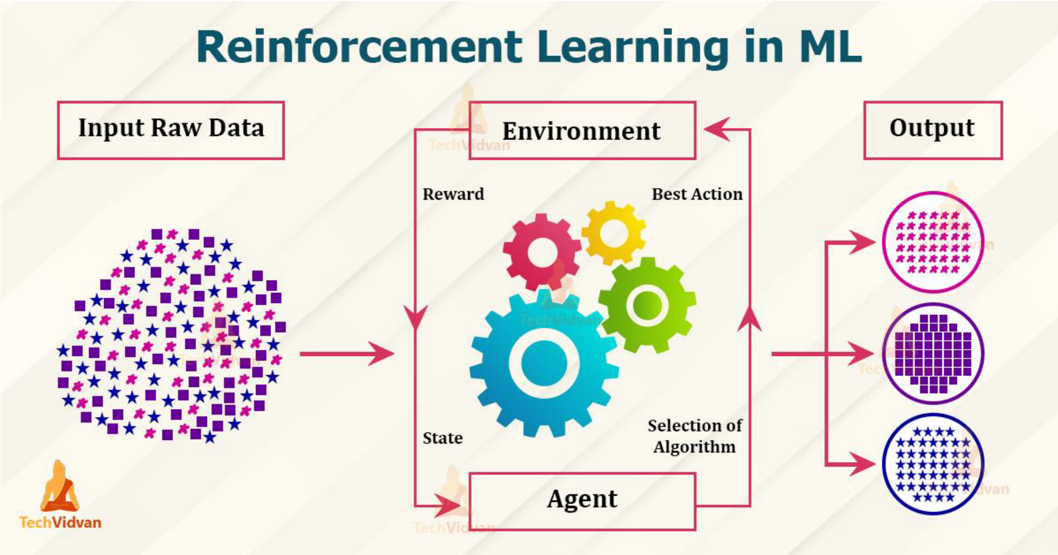

Figure 2.1: Reinforcement machine learning

learning model. These networks feature numerous hidden layers between the input and output layers, which is what gives them their depth and capability to handle complex expressions. DNNs are particularly effective for tasks that require learning complex patterns and representations from data, such as image recognition, natural language processing, and more.

2.3.1 Key Components of a DNN

Neurons (Nodes)

The fundamental computational components of a DNN are neurons, which accept the inputs in a similar fashion to how dendrites accept the input and through axons go to the neuron in biological neurons. After receiving input signals, each neuron applies an activation function, computes the weighted sum of the inputs, and generates an output. By adding non-linearity, the activation function enables the network to recognize intricate relationships in the data.

Layers

A DNN is organized into layers-wise structure to handle complex patterns. There are three main categories of layers:

- Input Layer: The input layer receives the first set of features or data to be processed.
- Hidden Layers: The layers between the input and output layers are called hidden layers. The network learns hierarchical data representations from the information flowing through hidden layers. These hidden layers define the capability of the neural network by increasing or decreasing the number of neurons in the hidden layers and itself the number of hidden layers.
- Output Layer: This layer generates the final prediction, classification, or regression result.

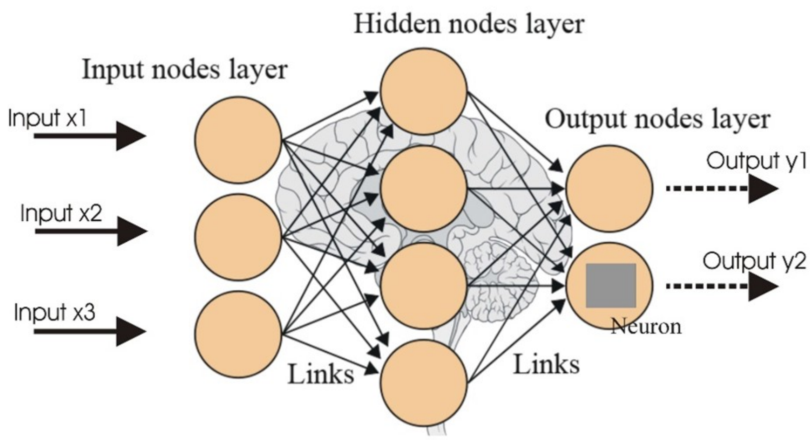

Figure 2.2: Deep Neural Network

Weights and Biases

The strength of a connection between neurons is determined by the weight assigned to it. Higher weight means higher importance, and lower weight means providing less importance to the input variable. A bias term is another feature of every neuron that aids in changing the activation function. During training, an optimization algorithm—typically backpropagation—teaches the network the ideal weights and biases. The optimization algorithm adjusts the weights and biases in such a way as to minimize the cost function(s), indicating that the model learns the dataset well.

Activation Functions

Activation functions give the network non-linearity, which enables it to efficiently approximate complicated functions. The sigmoid, ReLU (Rectified Linear Unit), PReLU (Parametric Rectified Linear Unit), and tanh (hyperbolic tangent) are examples of common activation functions.

- **ReLU**-The Rectified Linear Unit (ReLU) activation function replaces negative input values with zero, effectively causing neurons to be inactive when the input is below zero. This simple yet effective function helps mitigate the vanishing gradient problem and accelerates convergence during training. [**?**]

$$\mathrm{f(x)} = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

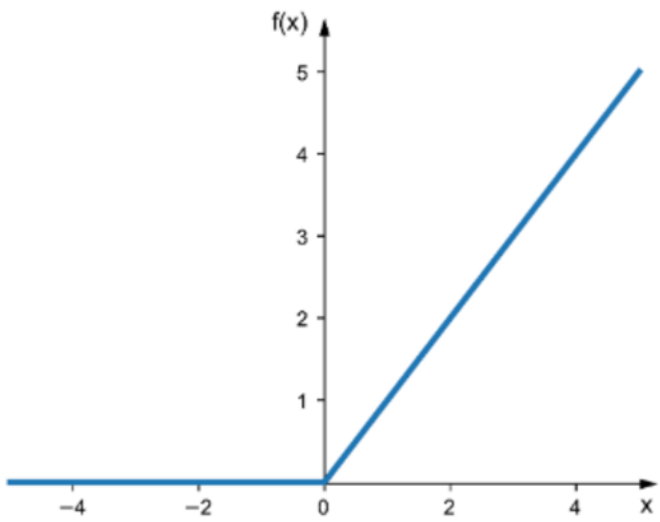

Figure 2.3: ReLU-function

- **Sigmoid**- The sigmoid activation function, also known as the logistic function, squashes input values to a range between 0 and 1. It is often used in binary classification tasks, where the output represents probabilities of belonging to a certain class. However, it suffers from the vanishing gradient problem and is prone to saturation for extreme input values, leading to slower training and potential gradient explosion.

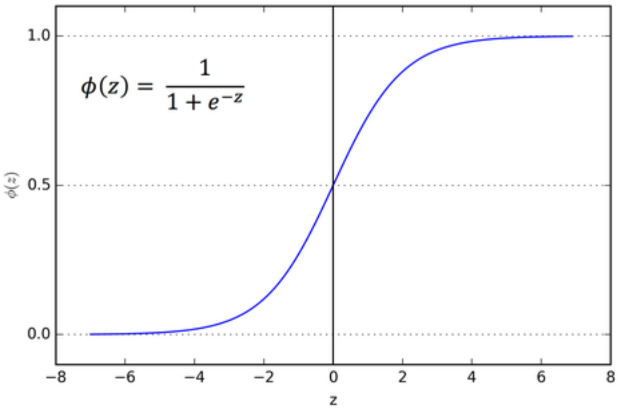

Figure 2.4: Sigmoid Activation Function

The sigmoid activation function is defined as:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2.1}$$

Feedforward Propagation

Data moves across the network from the input layer to the output layer when feedforward propagation is used. Based on the activation function and the weighted sum of inputs, each neuron calculates its output and passes it to the subsequent layers up to the output layer.

Back Propagation

Backpropagation is a supervised learning algorithm crucial for training artificial neural networks. It involves a forward pass to compute network predictions, followed by error calculation comparing predictions to actual targets. The backward pass computes gradients of the loss function with respect to each network parameter. These gradients guide weight and bias updates through optimization algorithms like gradient descent, minimizing the loss and improving network performance. This iterative process continues until convergence, allowing neural networks to learn complex patterns from labeled data, making backpropagation a fundamental tool in various domains, including image recognition, natural language processing, and autonomous driving.

Forward Pass:

- Linear transformation:

$$z^{(l)} = W^{(l)} \cdot a^{(l-1)} + b^{(l)} \tag{2.2}$$

  where W is Weight vector and a is Input vector and b is Bias vector.

- Activation function:

$$a^{(l)} = \sigma(z^{(l)}) \tag{2.3}$$

Error Calculation:

$$J = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)}) \tag{2.4}$$

Where L, denotes the loss function and y is actual output vector and y hat is expected output vector.

Backward Pass:

- Gradient of the loss function with respect to the output layer:

$$\frac{\partial J}{\partial z^{(L)}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z^{(L)}} \tag{2.5}$$

- Gradient of the loss function with respect to the weights and biases:

$$\frac{\partial J}{\partial W^{(l)}} = \frac{1}{m}\frac{\partial J}{\partial z^{(l)}} \cdot a^{(l-1)T} \tag{2.6}$$

$$\frac{\partial J}{\partial b^{(l)}} = \frac{1}{m}\sum_{i=1}^{m}\frac{\partial J}{\partial z^{(l)}} \tag{2.7}$$

Gradient Descent:

$$W^{(l)} = W^{(l)} - \alpha\frac{\partial J}{\partial W^{(l)}} \tag{2.8}$$

$$b^{(l)} = b^{(l)} - \alpha\frac{\partial J}{\partial b^{(l)}} \tag{2.9}$$

Above both equations (1.8) and (1.9) is weight and bias updation equation.

2.3.2 Optimizer

An optimizer is a crucial component in training neural networks, determining how parameters are adjusted to minimize the loss function. Popular optimizers like stochastic gradient descent (SGD), Adam, and RMSprop iteratively update weights based on gradients, influencing the model's convergence speed and performance in machine learning tasks.

Adam

Adam, short for Adaptive Moment Estimation, is a popular optimization algorithm for training neural networks. It combines the benefits of both momentum and adaptive learning rates. Adam dynamically adjusts learning rates for each parameter, offering fast convergence and robust performance across a wide range of deep learning tasks.

$$\begin{aligned}
m_t &= \beta_1 m_{t-1} + (1-\beta_1) g_t \\
v_t &= \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \\
\hat{m}_t &= \frac{m_t}{1-\beta_1^t} \\
\hat{v}_t &= \frac{v_t}{1-\beta_2^t} \\
\theta_{t+1} &= \theta_t - \frac{\alpha}{\sqrt{\hat{v}_t}+\epsilon}\hat{m}_t
\end{aligned}$$

Here m_t and v_t represent the first and second moments of the gradients, respectively. $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected estimates of these moments. θ_t denotes the parameters being optimized, g_t represents the gradients, and α is the learning rate. β_1 and β_2 are the exponential decay rates for the first and second moments, respectively, and ϵ is a small constant to prevent division by zero. Adjust parameters accordingly based on your specific implementation.

Deep Architecture

The depth of a DNN, which refers to the number of hidden layers, is a key factor in its ability to model complex relationships in data. Deeper networks are capable of capturing more abstract and hierarchical features. An artificial neural network having several layers of connected artificial neurons is known as a Deep Neural Network (DNN), also sometimes referred to as a deep learning model.These networks feature numerous hidden layers between the input and output layers, which is what gives them their depth. DNNs are particularly effective for tasks that require learning complex patterns and representations from data, such as image recognition, natural language processing, and more .

2.3.3 Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a deep learning architecture tailored for handling structured grid data, such as images. CNNs excel in tasks related to image recognition, classification, and analysis, forming a foundational technology in the field of computer vision. These networks are designed to automatically and adaptively learn spatial hierarchies of features from input images, making them highly effective for visual data processing.

In essence, CNNs utilize layers of convolutional filters that capture local patterns and hierarchies in data, which are then processed through pooling and fully connected layers to achieve tasks like image classification and object detection. This ability to automatically and efficiently learn complex patterns makes CNNs a powerful tool in various image-related applications.

Key Components of CNNs:

Convolutional Layers: Extract features from the input image using convolutional operations with multiple filters. Pooling Layers: Reduce the dimensionality of feature maps, retaining essential information while reducing computational complexity.

Fully Connected Layers: Combine extracted features to make predictions or classifications.

Using Convolutional Neural Networks (CNNs) on Tabular Data

While CNNs are primarily designed for image and spatial data, they can also be applied to tabular data, although this is less common and usually not the first choice for such data types. However, some research and applications have demonstrated the potential benefits of using CNNs on tabular datasets, especially when the data exhibits some spatial or sequential structure.

Advantages of Using CNNs on Tabular Data

(1) Capturing Local Patterns: CNNs can capture local relationships and interactions between features, which can be particularly useful in tabular data with spatial or temporal dependencies.
(2) Feature Engineering: The convolutional layers can automatically learn useful feature representations, reducing the need for manual feature engineering.
(3) Dimensionality Reduction: Pooling layers in CNNs can help reduce the dimensionality of the data, making the model more efficient and potentially improving performance.

2.3.4 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) designed to learn from sequences of data over time. LSTMs address the limitations of traditional RNNs, particularly their difficulty in learning long-term dependencies due to the vanishing gradient problem.

Key Features of LSTMs:

(1) Memory Cells:
LSTMs contain special units called memory cells that can maintain information in memory for long periods, making them suitable for tasks where context and order matter.
(2) Gates: Forget Gate: Decides what information to discard from the cell state. Input Gate: Determines what new information to store in the cell state. Output Gate: Controls the output flow of information from the cell state.
These gates regulate the flow of information into and out of the cell state,

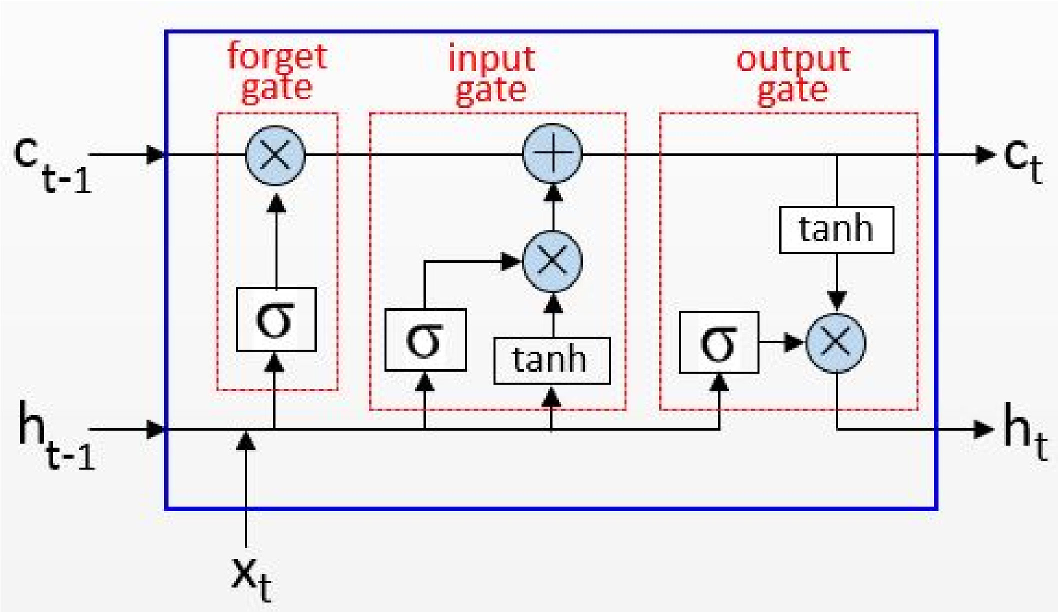

Figure 2.5: LSTM Network

allowing the network to retain and utilize relevant information over longer sequences effectively.

How LSTMs Work:

Forget Gate: The forget gate uses a sigmoid activation function to decide which information should be discarded from the cell state. It outputs a number between 0 and 1 for each number in the cell state C_{t-1}, where 1 means "completely keep this" and 0 means completely forget this.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Input Gate: The input gate updates the cell state with new information. It has two parts: a sigmoid layer that decides which values to update, and a tanh layer that creates a vector of new candidate values, $\tilde{C}_t$, that could be added to the state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Updating Cell State: The cell state is updated by combining the previous cell state and the new candidate values, weighted by the input gate.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Output Gate: The output gate determines the next hidden state. It decides which parts of the cell state to output.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

$$h_t = o_t * \tanh(C_t)$$

Advantages of LSTMs:

Handling Long-Term Dependencies: LSTMs can capture long-term dependencies in data, which is a significant improvement over traditional RNNs.
Better Gradient Flow: The gating mechanisms help maintain a better gradient flow during backpropagation, reducing issues like vanishing or exploding gradients.

2.4 Deep Learning and Photonics

Deep learning is a subset of machine learning that involves neural networks with multiple layers (deep neural networks). These networks, inspired by the human brain, can automatically learn hierarchical representations from data. Deep learning excels in tasks such as image and speech recognition, natural language processing, and pattern detection. Training deep neural networks often requires large datasets and substantial computational resources. The deep learning architecture allows the model to automatically extract relevant features from raw data, eliminating the need for manual feature engineering. Popular deep learning frameworks, such as TensorFlow and PyTorch, facilitate the development and training of complex neural network architectures.

Deep learning has significantly advanced the state-of-the-art in various domains, including computer vision, speech processing, and autonomous systems. The success of deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has fueled its widespread adoption in artificial intelligence applications. Ongoing research in deep learning focuses on improving model interpretability, addressing ethical considerations, and advancing techniques for efficient training.

2.4.1 Deep learning for forward nanophotonic modelling

In essence, forward modelling in nanophotonics consists in predicting the optical properties of photonic structures featuring subwavelength-scale complex features. Deep learning forward modeling in nanophotonics involves using neural networks to predict the optical response of nanoscale structures. These models are trained on datasets containing input parameters, such as nanostructure geometry, and corresponding optical properties. The trained networks efficiently capture the complex, nonlinear relationships between structure and optical behavior. This approach accelerates the simulation of light-matter interactions, aiding in the rapid and accurate prediction of nanophotonic device performance. Deep learning forward modeling is particularly valuable for its ability to automate and streamline the design and optimization of advanced nanophotonic structures .

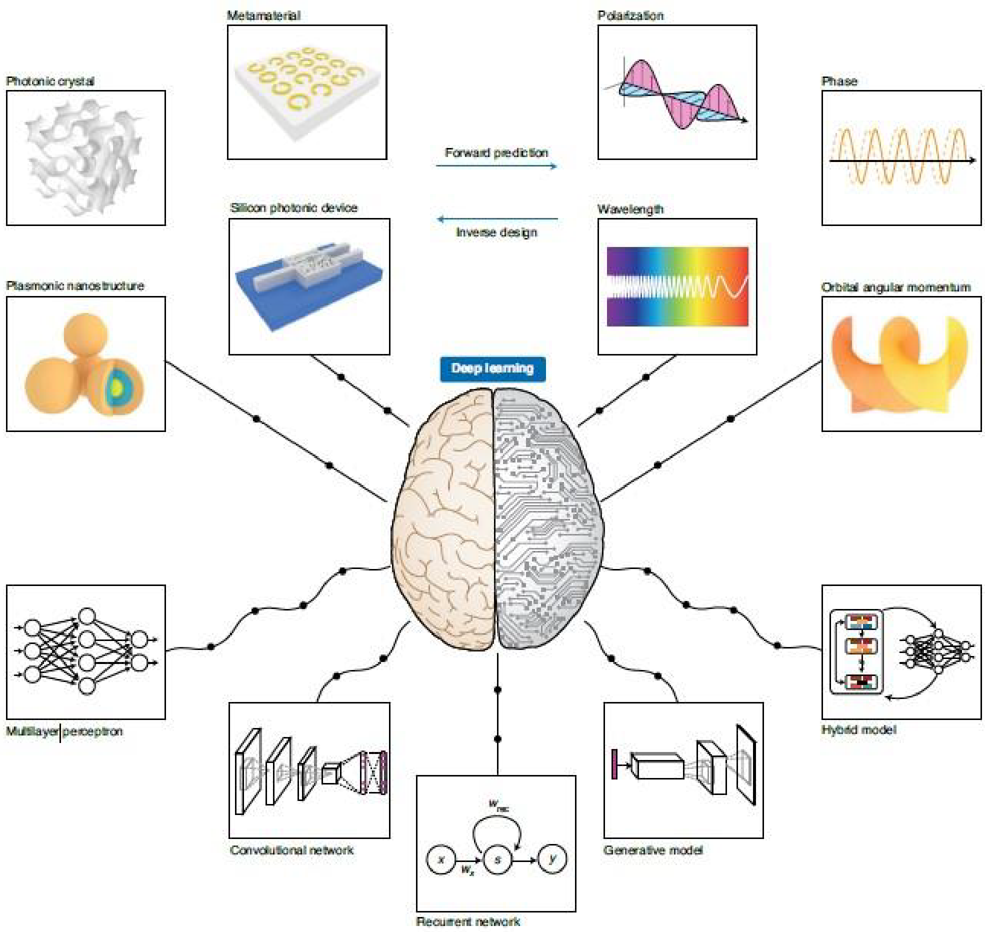

Figure 2.6: Application of DNN in Various Field [1]

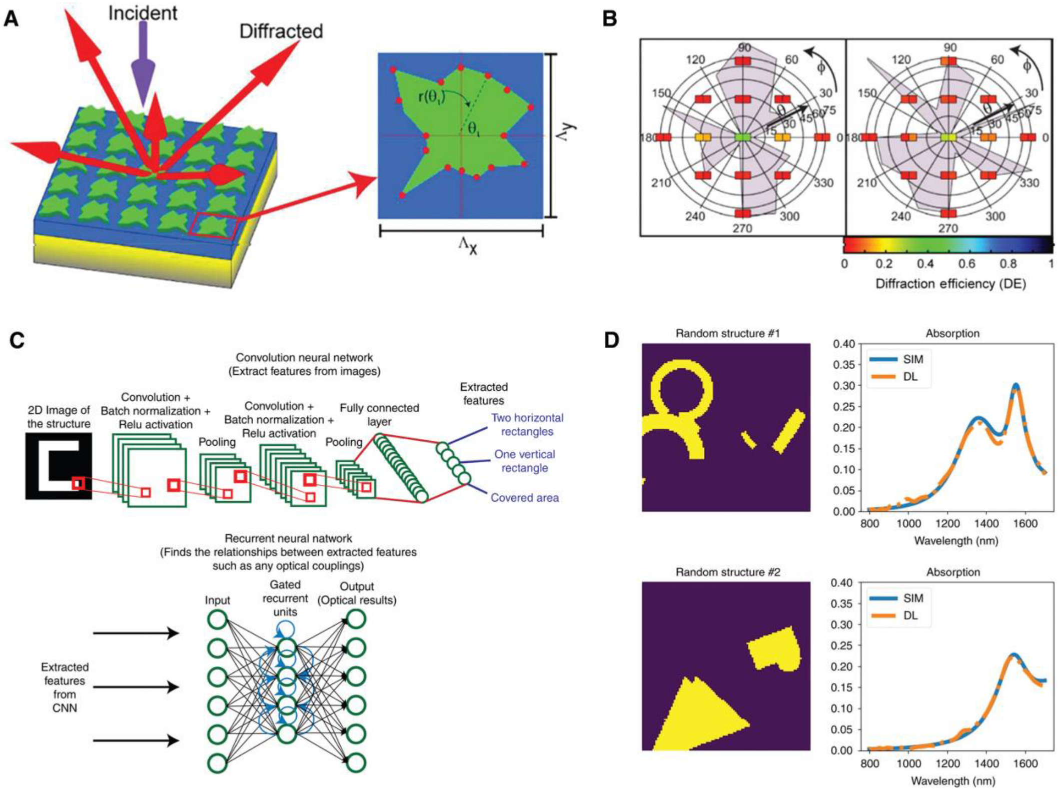

Figure 2.7: Forward Design [2]

2.4.2 Deep learning nanophotonic inverse design

Deep learning inverse design modeling in nanophotonics revolutionizes the process of creating optimal nanostructures by automating the exploration of design spaces. Neural networks are trained to predict nanophotonic structure geometries that yield specific desired optical properties. This approach enables the rapid discovery of novel and efficient structures without relying on traditional trial-and-error methods. By learning complex relationships between input parameters and desired outcomes, deep learning facilitates systematic and efficient optimization, making it a powerful tool for inverse design challenges in nanophotonics. The synergy of deep learning and nanophotonics accelerates the development of advanced optical devices and materials [2].

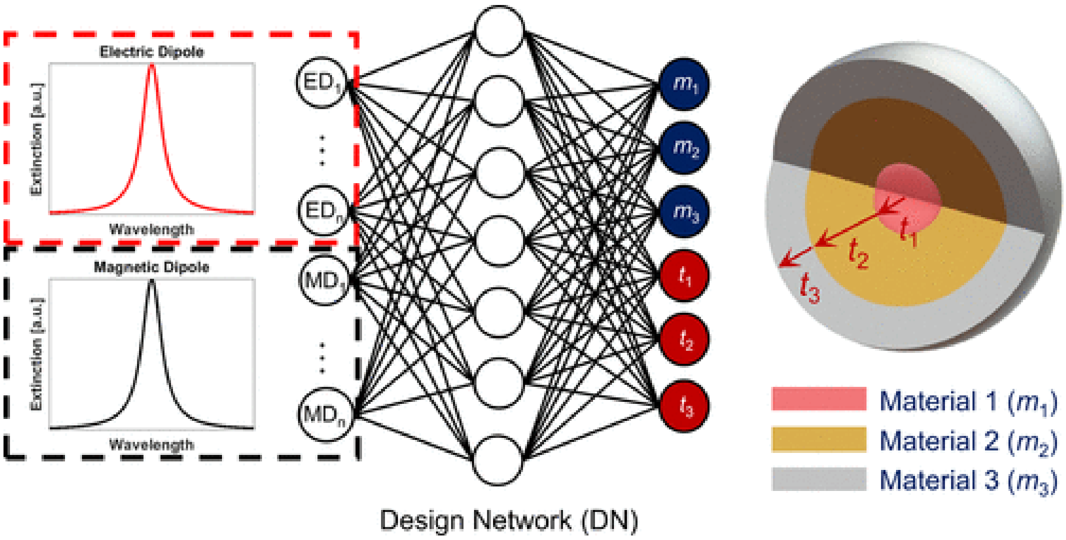

Figure 2.8: Inverse Design

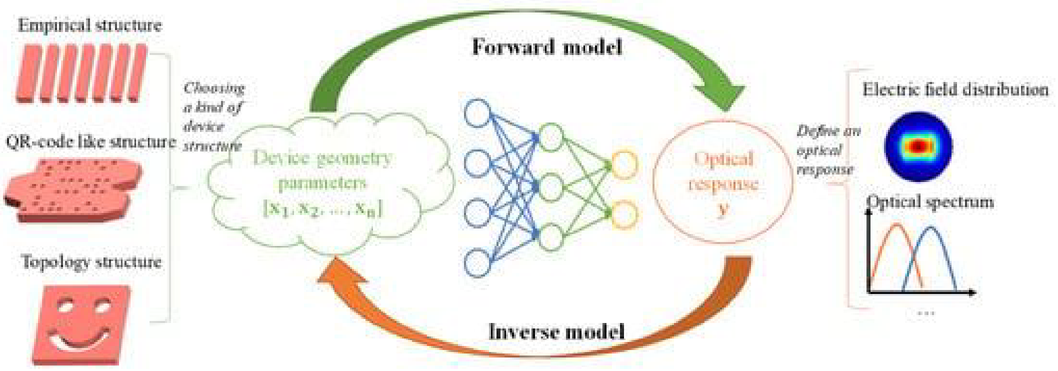

Figure 2.9: Nanophotonics Forward and Inverse Design

Chapter 3

Motivation and Problem Statement

3.1 Solar Cell

Introduction to Solar Cells

Solar cells, or photovoltaic cells, convert sunlight directly into electricity. They are made from semiconductor materials, primarily silicon, which exhibit the photovoltaic effect. When photons from sunlight strike the semiconductor, they excite electrons, creating electron-hole pairs. These charge carriers are then separated by an internal electric field, generating a flow of electric current.

The Need for Antireflection Coating

A significant challenge for solar cells is the reflection of incident light off their surface. Without an antireflection (AR) coating, a substantial portion of the light (up to 30%) can be reflected away, reducing the amount of light available for conversion into electricity. This reflection loss directly impacts the overall efficiency of the solar cell.

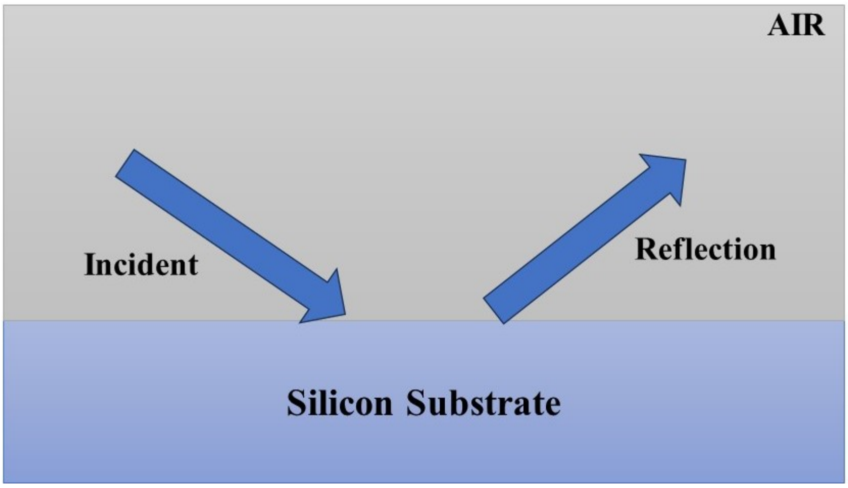

Figure 3.1: Solar Cell

Principles of Antireflection Coating

AR coatings are thin film layers applied to the surface of solar cells to minimize reflection and maximize light absorption. They work on the principle of destructive interference. When light waves reflect off the top and bottom surfaces of the AR coating, they can interfere destructively, canceling each other out and thereby reducing the overall reflected light. Materials Used in AR Coatings

Several materials are commonly used for AR coatings in solar cells, chosen for their optical properties and compatibility with the semiconductor material:

(1) Magnesium Fluoride (MgF2): Low refractive index, often used for single-layer coatings.

(2) Silicon Dioxide (SiO2): Durable and effective, commonly used in multilayer coatings.

(3) Titanium Dioxide (TiO2): High refractive index, often used in combination with other materials.

(4) Tantalum Pentoxide (Ta2O5): High refractive index, good for multilayer coatings.

(5) Silicon Nitride (Si3N4): Provides excellent passivation and AR properties, commonly used in silicon solar cells.

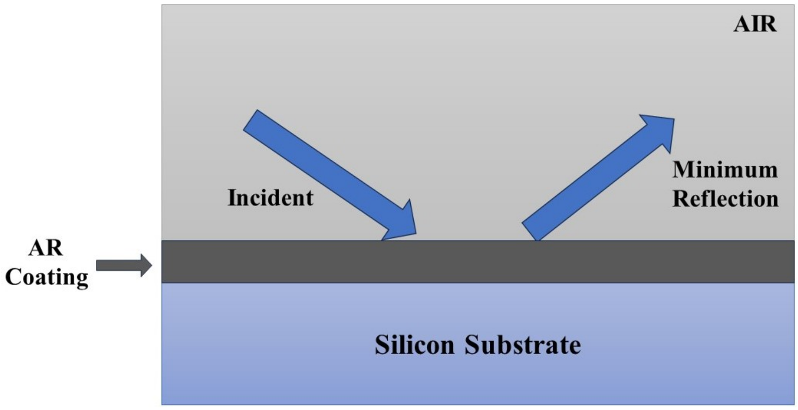

Figure 3.2: Solar cell With AR Coating

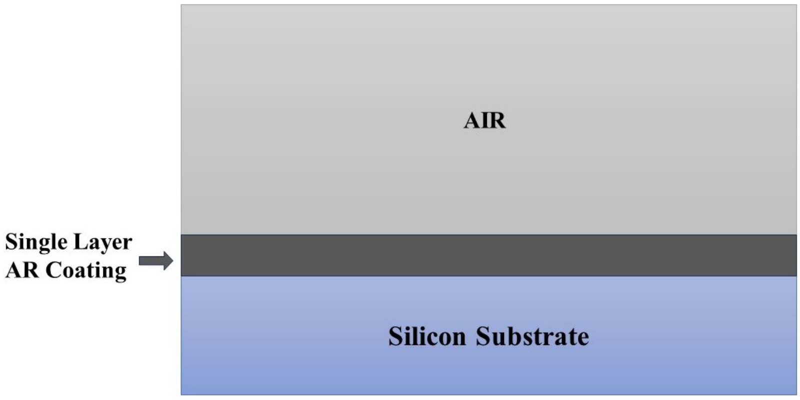

Figure 3.3: Single Layer AR Coating

3.1.1 Types of AR Coatings

Single-Layer Coatings: These consist of a single material layer, typically with an optical thickness of a quarter wavelength $\frac{\lambda}{4}$ of the target light wavelength. They are simple to apply and provide moderate reduction in reflection.
Multi-Layer Coatings: Comprise multiple layers of different materials, each designed to target different wavelengths of light. These coatings can achieve very low reflectance over a wide spectrum, significantly improving efficiency.

Design Considerations for AR Coatings

Thickness: The thickness of each layer in an AR coating is critical for achieving the desired interference effect. It is typically optimized to be a quarter of the wavelength of the incident light within the material.
Refractive Index Matching: The refractive indices of the coating materi-

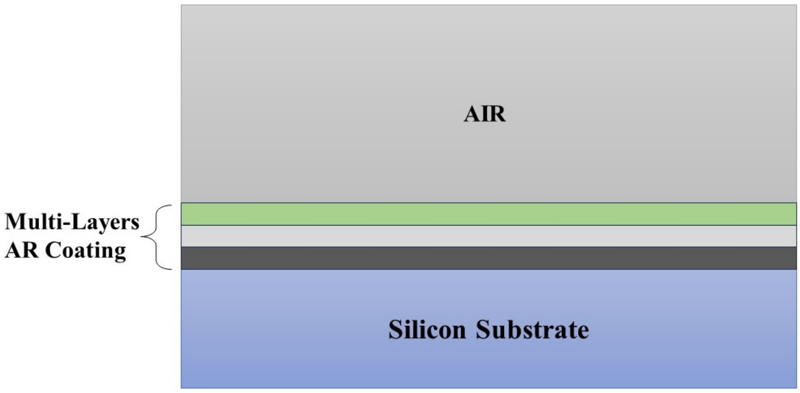

Figure 3.4: Multi-Layer AR Coating

als are chosen to create the necessary destructive interference. The goal is to gradually transition the refractive index from air to the solar cell surface.

Durability: AR coatings must withstand environmental conditions such as UV radiation, temperature changes, and humidity, without degrading in performance.

Impact on Solar Cell Efficiency

By reducing the reflection of incident light, AR coatings allow more light to enter the solar cell, increasing the number of photons available to generate electron-hole pairs. This increase in absorbed light leads to a higher generation of electric current and, consequently, a higher efficiency of the solar cell. Specifically, AR coatings can reduce reflection losses from around 30% to less than 5%, significantly boosting the overall energy conversion efficiency.

3.2 Problem Statement

The efficiency of solar cells is significantly impacted by the reflection losses at their surface. To mitigate these losses, multilayer antireflection (AR) coatings are applied, which consist of various materials and thicknesses. However, designing optimal AR coatings that maximize solar cell efficiency is a complex and time-consuming process, requiring extensive trial and error to determine the best combination of materials and their corresponding thicknesses.

Current approaches in AR coating design rely heavily on empirical methods and simulations, which are not only resource-intensive but also limited in their ability to explore the vast parameter space of potential designs. This leads to suboptimal solutions and limits the advancements in solar cell technology.
Research Objective
This research aims to address the challenges in designing multilayer AR coatings for solar cells by leveraging advanced machine learning techniques, specifically Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks. The primary objectives are:

(1) Predicting Layer Thickness: Develop a machine learning model capable of accurately predicting the thickness of each layer in a multilayer AR coating based on given input parameters.
(2) Material Selection: Utilize the machine learning model to select the most suitable material for each layer from a predefined list of materials, optimizing the overall AR coating design for maximum efficiency.

3.2.1 Research Significance

By integrating machine learning models into the design process of AR coatings, this research aims to significantly reduce the time and resources required for optimization, enabling more efficient and effective exploration of potential coating designs. This will not only enhance the efficiency of solar cells but also contribute to the broader adoption of solar energy by making it more cost-effective and competitive with traditional energy sources.

3.2.2 Expected Outcomes

Accurate Predictive Models: The development of robust DNN, CNN, and LSTM models that can accurately predict the optimal thickness and material for each layer in a multilayer AR coating.
Enhanced Solar Cell Efficiency: Achieve higher solar cell efficiencies through optimized AR coatings designed using the developed machine learning models.

Streamlined Design Process: Provide a framework for the solar cell industry to adopt machine learning in the AR coating design process, leading to faster and more efficient innovation cycles.

This research will contribute to the field of nanophotonics and solar energy by demonstrating the potential of machine learning to revolutionize the design of key components in solar cell technology.

Chapter 4

Proposed Architecture and Implementation

4.1 Python Libraries

Python offers a rich ecosystem of libraries that are pivotal for data science, machine learning, and scientific computing. Some of the most widely used libraries include TensorFlow, Scikit-learn, Pandas, and Matplotlib. Each of these libraries serves a specific purpose and greatly simplifies the process of data analysis, model building, and visualization.

4.1.1 TensorFlow

TensorFlow is an open-source machine learning library developed by Google. It provides a comprehensive ecosystem for developing and deploying machine learning models, from research prototypes to production systems. TensorFlow supports both deep learning and traditional machine learning algorithms, making it versatile for a wide range of applications. Its high-level API, Keras, allows for quick prototyping and easy model experimentation.

4.1.2 Scikit-learn

Scikit-learn, also known as Sklearn, is a powerful library for classical machine learning built on top of NumPy, SciPy, and Matplotlib. It provides simple and efficient tools for data mining, data analysis, and machine learning. Scikit-learn supports a wide variety of supervised and unsupervised learning algorithms, including regression, classification, clustering, and dimensionality reduction, making it a go-to library for many machine learning practitioners.

4.1.3 Pandas

Pandas is an essential data manipulation and analysis library for Python. It provides data structures like DataFrames and Series, which allow for fast and flexible data manipulation. Pandas is particularly useful for handling structured data, performing operations like merging, reshaping, aggregating, and cleaning datasets. Its intuitive syntax and rich functionality make it a staple for data analysts and scientists.

4.1.4 Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. It is highly customizable and capable of producing a wide range of plots and charts, from simple line plots to complex 3D visualizations. Matplotlib integrates well with other Python libraries, such as Pandas and NumPy, enabling seamless data visualization workflows. Its ability to create publication-quality graphics makes it a popular choice for researchers and developers alike.

4.2 ML Model 1

4.2.1 DNN

A Deep Neural Network (DNN) is a type of artificial neural network with multiple hidden layers between the input and output layers, enabling it to model

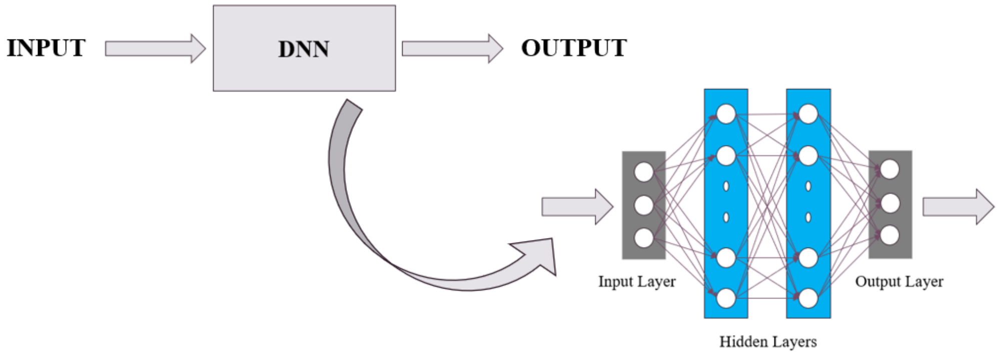

Figure 4.1: DNN Network for both Thickness and Material

complex, non-linear relationships. Unlike Multi-Layer Perceptrons (MLPs), which typically have one or two hidden layers, DNNs can have many layers, making them capable of learning hierarchical feature representations. This deeper architecture allows DNNs to excel in tasks like image and speech recognition. While MLPs are suitable for simpler problems, DNNs are used for more complex and large-scale data applications due to their greater depth and capacity.

4.3 ML Model 2

This machine learning model is designed for predicting thickness and material in multilayer thin film antireflection coatings without using a feedback mechanism in the LSTM network. It follows a standard procedure of loading and preprocessing data from a CSV file, then splitting it into training and testing sets for features, thickness, and material labels, and reshaping the input data for compatibility with neural network layers.

The model architecture consists of an input layer followed by several 1D convolutional layers (Conv1D) to extract features from the input data. These features are then fed into a two-layer LSTM network without feedback, which aims to capture temporal dependencies in the data.

For thickness prediction, a dense neural network (MLP) is constructed on top of the LSTM outputs, comprising three dense layers and a final output layer

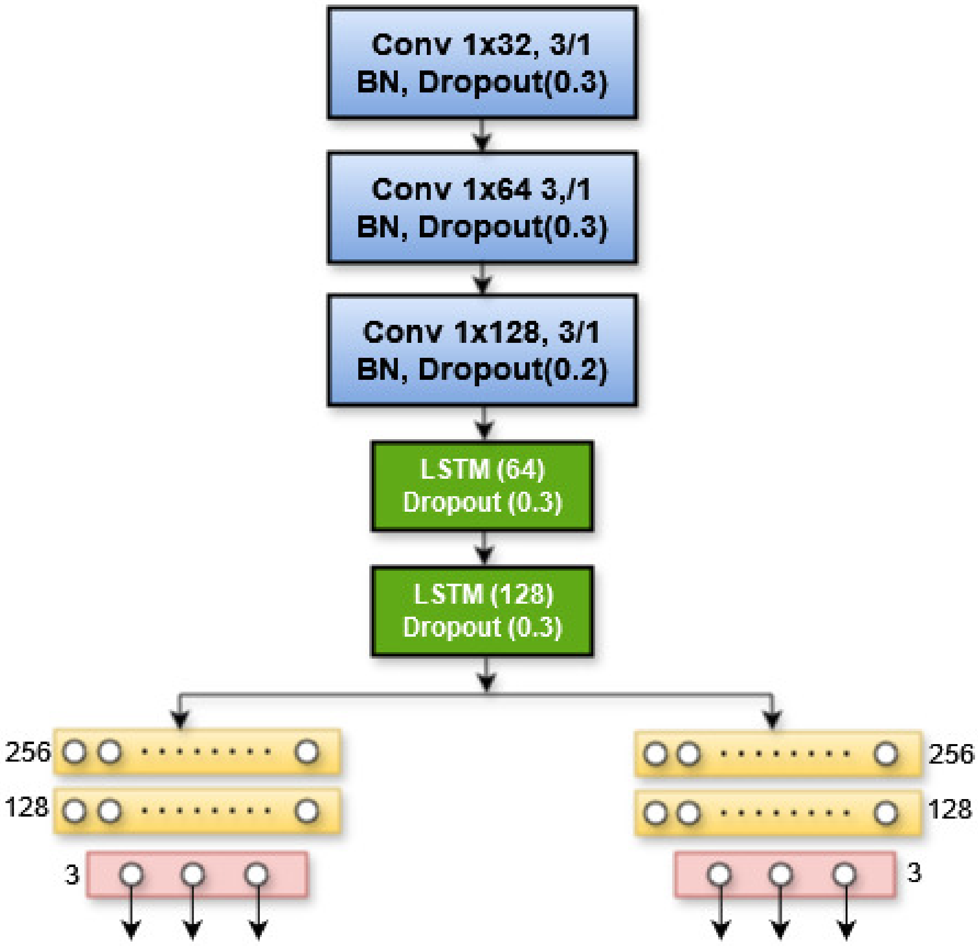

Figure 4.2: Model (LSTM without Feedback)

with three units for each thickness value. The model is compiled using the Mean Squared Error loss function and trained on the thickness data.

For material prediction, the LSTM output is connected to another dense network with three separate output layers. Each layer predicts the material for a different layer using softmax activation for multi-class classification. The material prediction model is compiled with categorical cross-entropy loss and accuracy metrics and trained on the corresponding material data.

The model's performance is evaluated on the test set, providing mean squared error for thickness predictions and accuracy for material predictions. This approach, while similar to models with feedback mechanisms in LSTMs, allows for a comparison in modeling capabilities for predicting thickness and material in multilayer thin film antireflection coatings

4.4 ML Model 3

This machine learning model is designed for thickness and material prediction in multilayer thin film antireflection coatings, incorporating a feedback mechanism in the LSTM network. It starts by loading and preprocessing data from a CSV file, splitting it into training and testing sets for features, thickness, and material labels. The input data is reshaped for compatibility with the neural network layers.

The model architecture includes an input layer followed by multiple 1D convolutional layers (Conv1D) to extract features. These features are then fed into a two-layer LSTM network with a feedback mechanism to capture temporal dependencies and generate a rich feature representation.

For thickness prediction, a dense neural network (MLP) is built on top of the LSTM outputs, consisting of three dense layers and a final output layer with three units for each thickness value. This sub-model is compiled using the Mean Squared Error loss function and trained on the thickness data.

For material prediction, the LSTM output is connected to another dense network with three separate output layers, each predicting the material for a different layer using softmax activation for multi-class classification. The material prediction model is compiled with categorical cross-entropy loss and accuracy metrics, and trained on the corresponding material data.

The model's performance is evaluated on the test set, providing mean squared error for thickness predictions and accuracy for material predictions. This approach leverages both convolutional and recurrent neural network techniques, along with the feedback mechanism in the LSTM, to effectively model complex dependencies in multilayer thin film coatings.

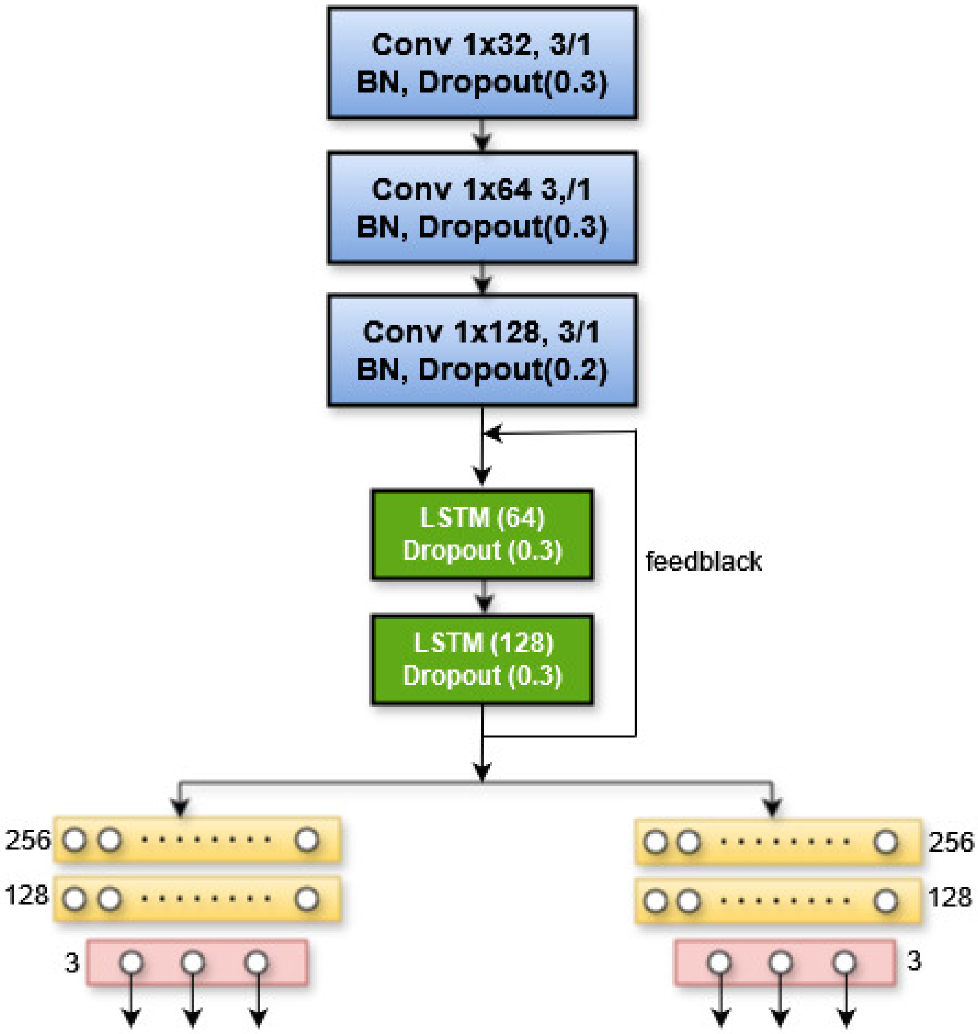

Figure 4.3: Model (LSTM with Feedback)

Chapter 5

Results and Conclusion

The goal of this study was to achieve a reflectance value between 0 and 1 percent across the wavelength range of 400nm to 800nm for multilayer thin film antireflection coatings. The results were obtained using three different machine learning models for thickness and material prediction, evaluated at various epochs. The models included a convolutional neural network (CNN) for feature extraction, followed by a two-layer LSTM network for capturing temporal dependencies, and a dense neural network (MLP) for thickness and material prediction. The models were trained and tested on a dataset of features, thickness, and material labels, and their performance was assessed based on mean squared error for thickness predictions and accuracy for material predictions. The results indicate that the models achieved a reflectance value within the desired range, demonstrating their effectiveness in designing antireflection coatings for solar cells [6].

5.1 Predictions Through Models

5.2 Conclusion

In conclusion, this study aimed to design and compare three different models for thickness and material prediction in multilayer thin film antireflection coatings. The models were based on different neural network architectures,

#MODEL_1 Prediction Epoch-10

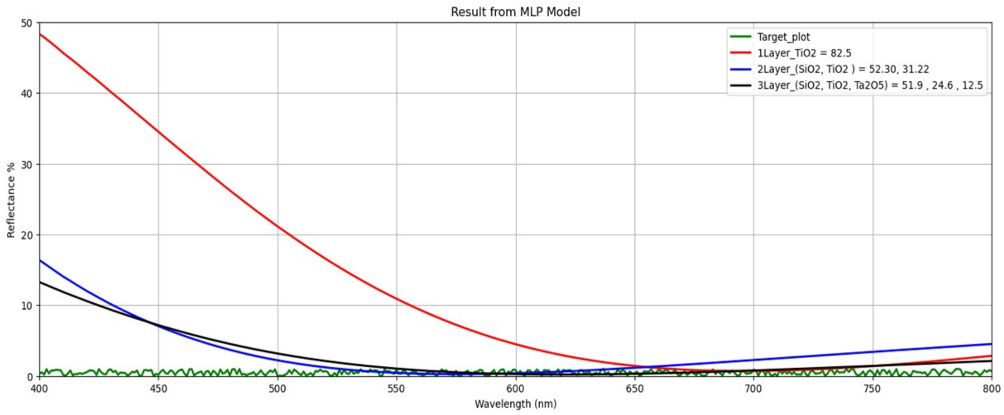

Figure 5.1: Epoch 10

#MODEL_1 Prediction Epoch-50

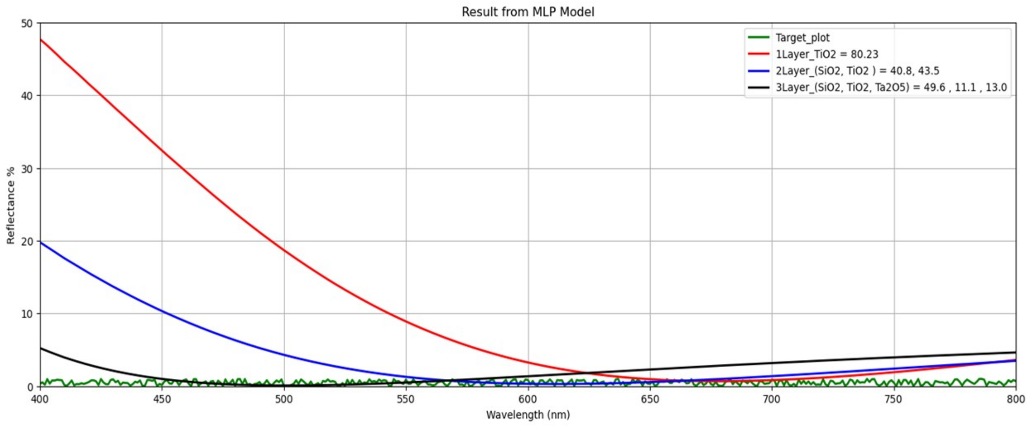

Figure 5.2: Epoch 50

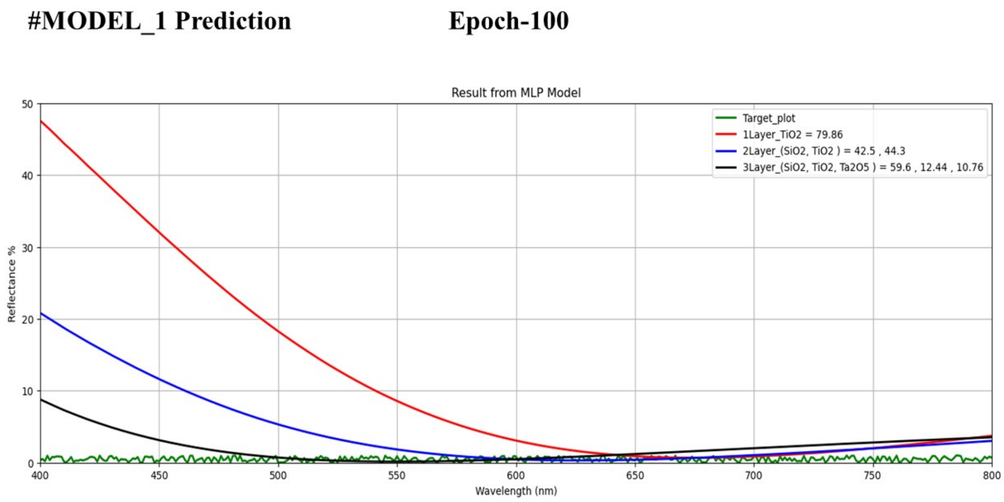

Figure 5.3: Epoch 100

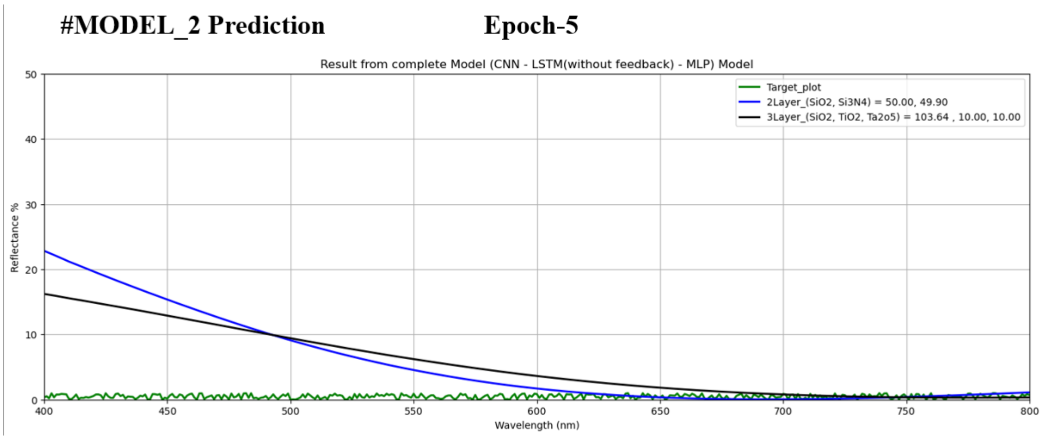

Figure 5.4: Epoch 5 (LSTM without Feedback)

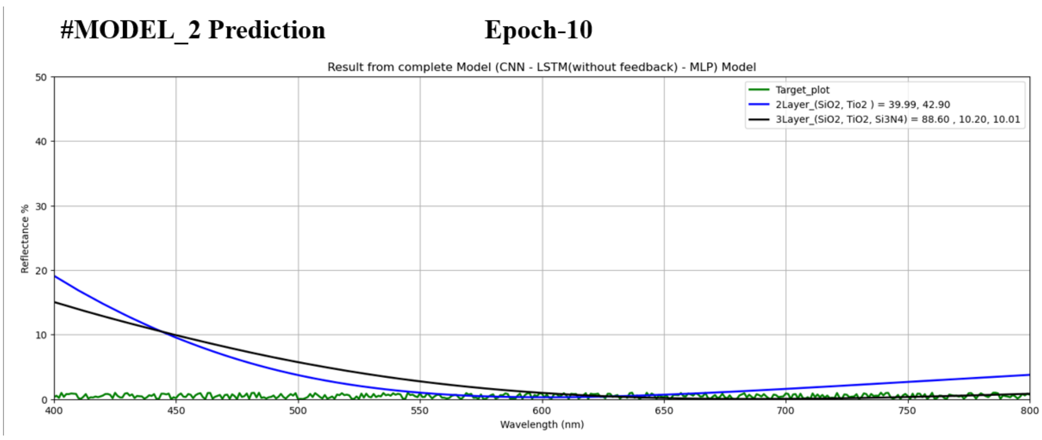

Figure 5.5: Epoch 10 (LSTM without Feedback)

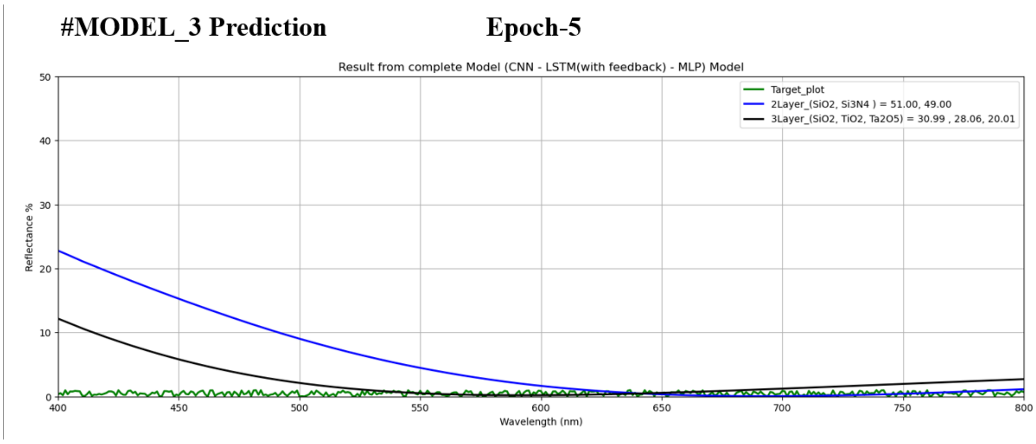

Figure 5.6: Epoch 5 (LSTM with Feedback)

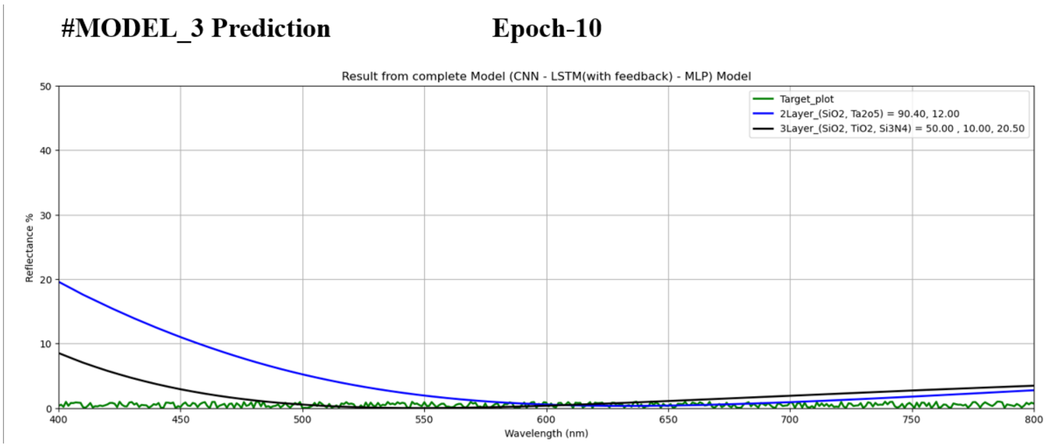

Figure 5.7: Epoch 10 (LSTM with Feedback)

including DNN, MLPs, and LSTM networks.

After training and testing the models, it was found that model 2 without feedback and model 3 with feedback performed better in achieving the target plot of reflectance between 0 and 1 percent across the wavelength range of 400nm to 800nm. These models demonstrated superior performance in covering a wider range of wavelengths compared to the DNN-based model.

The results suggest that incorporating feedback mechanisms, as in model 3, can enhance the model's ability to capture complex temporal dependencies and improve predictions for multilayer thin film coatings. This highlights the importance of model architecture in achieving accurate and efficient predictions for such applications.

5.3 Future Work

(1) Enhanced Feature Engineering: Investigate advanced feature engineering techniques to extract more informative features from the input data. This could involve exploring different representations of the coating layers or incorporating domain knowledge to improve model performance.
(2) Application to Other Coating Types: Extend the application of the models to other types of coatings beyond antireflection coatings. This could include exploring the use of the models for designing and optimizing coatings for different applications, such as optical filters, mirrors, or protective coatings, to broaden the impact of the research.

References

[1] W. Ma, Z. Liu, Z. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu, "Deep learning for the design of photonic structures," *Nature Photonics*, vol. 15, pp. 1–14, 10 2020.

[2] S. So, T. Badloe, J. Noh, J. Bravo-Abad, and J. Rho, "Deep learning enabled inverse design in nanophotonics," *Nanophotonics*, vol. 9, no. 5, pp. 1041–1057, 2020.

[3] S. J. Byrnes, "Multilayer optical calculations," 2020.

[4] A. Luce, A. Mahdavi, F. Marquardt, and H. Wankerl, "Tmm-fast, a transfer matrix computation package for multilayer thin-film optimization: tutorial," *J. Opt. Soc. Am. A*, vol. 39, pp. 1007–1013, Jun 2022.

[5] J. an qing, O. Yoshie, and L. Chen, "Multilayer optical thin film design with deep q learning," *Scientific Reports*, vol. 10, 07 2020.

[6] M. J. Hossain, D. Reitano, and A. S. Rakin, "Inverse design of silicon photonics components: A study from deep learning perspective," in *2023 IEEE Photonics Conference (IPC)*, pp. 1–2, 2023.