Logical Supervenience of Consciousness on the Physical & Prog Rock Music Classification Approaches and Techniques

By Apoorv Chandurkar, Shaileshbhai Gothi, Daniel Irizarry, Siddardha Majety, and Shandy Sulen

I. Introduction

Countless philosophers agree that most "things" are logically supervenient on the physical. However, to make this claim for consciousness is quite hotly controversial and welcomes numerous arguments that support or deny the question "*Is consciousness logically supervenient on the physical?*". In Section II, supervenience definitions and terminology are discussed that lay the groundwork for some of the arguments presented in Section III, which holds three separate claims that agree with the question, deny the question, and argue that the question at hand is irrelevant. Section IV ends the philosophical component of the paper with a discussion relating the question of logical supervenience of consciousness on the physical with present and future development of artificial intelligence.

Algorithms in the field of machine learning are typically partitioned into solving two types of problems: classification and regression. The problem proposed is to classify an audio signal as to whether it belongs to the progressive rock music genre. While this could be considered as a trivial binary classification problem, the complexity of the raw data, coupled with the loose definition of the progressive rock genre, bring forth difficulty in the classification process. In order to fully utilize the time series nature of audio files, a specific time-series classification model, long short-term memory (LSTM), was employed to deliver maximum performance. Section V details the set of features used to train the multitude of classification models that are discussed in Section VI along with the performance of the models on training and testing sets. Section VII provides an ending conclusion regarding the Progressive Rock music classification problem. Lastly, Section VIII contains the bibliography of all necessary sources.

II. Types of Supervenience

Within the realm of philosophy, supervenience is a relational notion between sets of properties used by philosophers to describe relationships between things (which often incites varying degrees of controversy among sects of philosophers). The widely accepted definition of supervenience states that a set of properties *B* supervenes on a set of properties *A* when no two things can vary with respect to *A* without also varying with respect to *B*. In simpler words, there cannot be a *B* difference without an *A* difference. When the "things" in question are individuals, local supervenience is satisfied. Likewise, when the "things" are worlds, global supervenience is satisfied. On a separate but closely related note, supervenience can hold with differing degrees of modal force, where logical supervenience is often treated as "strong" and natural supervenience as "weak".

A. Local & Global Supervenience

B-properties supervene locally on A-properties if the A-properties of an individual determine the B-properties of that individual. Consider the case of a $100 USD bill issued by the U.S. Federal Reserve and a counterfeit bill. The counterfeit bill has the exact same physical properties as the one issued by the U.S. Federal Reserve, i.e. the shape, color, weight, etc. between the two bills are identical. The property of shape is locally supervenient on the physical, just as the property of color or weight is locally supervenient on the physical. The physical defines the arrangement of each atom that makes up the bills; thus, the physical dictates the shape, color, and weight of the bills. However, a context-dependent property, like value, is not locally supervenient on the physical. Consider how the USD bill has a value of $100 (as viewed by a citizen of the United States of America) but the counterfeit bill has a value closer to $0. Even though both bills are physically identical, they still differ in value; thus, value does not locally supervene on the physical. Later, the claim will be made that value is dependent on consciousness, which explains its lack of supervenience on the physical. Local supervenience can be broken down further into weak and strong supervenience. Weak supervenience entails that there is no possible world that contains individuals that share identical A-properties with varied B-properties. Strong supervenience states that there are no possible individuals that share identical A-properties with varied B-properties whether the individuals are in the same world or different worlds. When the range of worlds is the same, strong supervenience propositions entail weak supervenience propositions, but not vice versa. That is, although local supervenience holds for the U.S. Federal Reserve bill and the counterfeit bill in our actual world, given several differently-defined worlds that hold these items, global supervenience may not be satisfied.

To understand global supervenience, let us first define how a world is understood in philosophy. A world is what humans perceive as a universe, which encompasses past, present, future, and every possible detail. B-properties supervene globally on A-properties if the A-properties of the world determine the B-properties of that world. For example, two worlds that are physically identical must also be biologically identical since the arrangement of every particle in physical space-time of a world dictates the biological structure of that world. Global supervenience is often utilized by philosophers to make a claim that is unable to be made with either strong or weak individual supervenience. Interestingly, although biological properties *globally* supervene on the physical, they may not *locally* supervene on the physical. Consider the example where two biological organisms within a world have differing evolutionary histories. One could then argue that local supervenience on the physical is not satisfied. Understandably, if two things are locally supervenient, then they are globally supervenient, but not vice versa as we see in this example.

B. Logical & Natural Supervenience

As mentioned earlier, supervenience can hold with varying degrees of modal force: logical, metaphysical, and natural (nomic). For the purposes of this paper, we assume that metaphysical necessity is just as strong as logical necessity. Also, note that the logical possibility or necessity referred to in this paper refers to "broadly logical" possibility. A set of B-properties supervenes logically on A-properties if there are no two logically possible situations that are identical with respect to A-properties but differing in B-properties. Logical possibilities are unconstrained by the laws of the natural world and include any worlds that are conceivable and non-contradictory. For example, a world with inedible food is contradictory – and thus logically impossible – as the very definition of food implies edibility. Yet, a world where dragons exist is a logical possibility since it can be conceived and is non-contradictory. When B-properties logically supervene on A-properties, we can say that the A facts entail the B-facts. Philosophically speaking, property *P* entails property *Q* when anything that possesses *P* also possesses *Q*; however, it's important to distinguish the fact that supervenience does not always suffice for entailment. With logical supervenience, the set of B-facts that supervene on the A-facts comes "for free", i.e. nothing extra must be done to achieve the set of B-facts. This idea plays a part in making logical supervenience a "strong" form of supervenience.

Whereas logical supervenience considers all logically possible words, natural (or nomic) supervenience is constrained to the natural laws of a defined, real world. A naturally possible situation must occur in nature without infringing upon any specified natural laws of the world. For example, an elephant piloting a plane on a trans-Atlantic flight is a natural possibility (even though it has yet to happen) since no natural laws are technically violated. A real-word example of natural supervenience is Newton's second law of motion which states that the force on an object is dependent upon the mass and acceleration of that object. Thus, it can be said that force is naturally supervenient upon the properties of mass and acceleration as the force of an object cannot change without the mass or acceleration of the object also changing. This example also shows that natural supervenience requires laws in place to relate the physical with the property that is supervening upon it. Put another way, the ontology, or the nature of being, regarding the natural supervenience of the set of B-properties does not immediately follow, given the set of A-properties. In philosophical terms, the set of B-properties are "over and above" the set of A-properties. The significant constraints that are placed within the realm of natural possibility cause natural supervenience to be viewed as a "weaker" relation than logical supervenience. Clearly, all natural possibilities in a particular world are logical possibilities as well, because natural possibilities are intrinsically conceivable and non-contradictory.

III. Opinions

To argue the claim that consciousness is or isn't logically supervenient on the physical, the tricky task of defining consciousness must first be performed. In "The Conscious Mind", David Chalmers distinguishes psychological consciousness and phenomenal consciousness. Psychological consciousness concerns itself with awareness, introspection, and attention among other things. Cognitive models and researchers can explain various facets of psychological consciousness. Phenomenal consciousness – however – concerns itself with the fundamental idea of experience and is significantly more complicated to understand and analyze. Phenomenal consciousness is the subject that is referred to in the question "*Is consciousness logically supervenient on the physical?*". Three separate opinions in the successive subsections are formed to support, deny, and refuse discussion altogether of the question.

A. Opinion 1: Consciousness is not logically supervenient on the physical

There is little to show that external indicators of consciousness exist in any other organism, a tenet of the philosophical concept of solipsism. As the notable Latin phrase goes "Cogito, ergo sum", which translates to "I think, therefore I am", yet there is absolutely no metric of evidence of consciousness in anything else. Yes, electrodes can be stuck to a human brain to identify regions of billions of neurons activating within the brain per some stimuli, but this does not come close to explaining phenomenal consciousness. It stands as an utmost challenge to identify consciousness in another human, much less any living organism. Consider that you live everyday surrounded by people where their individual instances of consciousness are never absolutely guaranteed. It is not so difficult then to conceive the logical possibility of a psychological zombie. To understand the meaning of a psychological zombie, the reader should imagine a doppelganger of himself or herself that reacts in the same way when getting bit by a mosquito, enjoying chocolate chip brownies, and holding conversations about Star Wars, along with all other conceivable acts. Yet, none of these activities or events are coupled by experience for the doppelganger. It is clear that the idea of a zombie is naturally impossible, i.e. given the laws of our natural world, an exact clone of the reader would also have consciousness since the biological structure in the two specimens are exactly the same. However, the logical possibility of a zombie is coherent, i.e. the conceivability of a psychological zombie is not impossible. Consider a world physically the same as our own where experiential consciousness does not exist. That is, the two sets of B-properties differ about an identical set of A-properties. It is this argument that suggests that consciousness does not logically supervene globally on the physical.

To continue the argument, a provision of reductive explanation is in order. Reduction states that if a set of B-properties reduces to a set of A-properties, then there cannot be a B-difference without an A-difference. This is true due to the reflexive nature of supervenience, which states that a set of A-properties trivially supervenes on a set of A-properties, which can be understood as it being impossible to have an *A* difference without an *A* difference. A set of B-properties logically supervenes on a set of A-properties, if a reductive explanation exists. For example, heat logically supervenes on the physical because heat can be reductively explained as the average kinetic energy of atoms in a space which can continuously be reductively explained in lower-level terms until the set of physical properties are reached. The reductive explanation for consciousness on the physical does not exist. The current state of knowledge of the biological structure of an organism or the arrangement of neurons in the brain, which are both logically supervenient on the physical, is not enough to reductively explain consciousness. Reductive explanation is also fundamentally particular, meaning that it accounts for only particular instances of a phenomenon. A plethora of reductive explanations of a particular truths would then create a general truth. Logical supervenience of consciousness on the physical fails if *any* kind of reductive explanation for *any* kind of consciousness fails, which complicates the task even more so. The lack of an accepted reductive explanation for the supervenience of consciousness on the physical points to consciousness being only naturally supervenient. If consciousness is naturally supervenient, an extra set of laws must be provided that "maps" the physical set of properties to the consciousness set of properties.

Consciousness is a remarkably unique concept as our primary evidence of its existence derives only from our own subjective case of it. Unlike nearly all other things, there lies an "epistemic asymmetry", or an imbalance of external knowledge, with regards to phenomenal consciousness. To understand and know the consciousness of another human being can only be inferred through external cues of that human being. A hot air balloon, rock, and robotic arm, while all being wildly different characteristically are similar in that they can all be equally and externally observed. A hot air balloon, rock, and robotic arm can all be reductively explained by the physical, like any other externally observable item. Even things that are intangible like the economy or community can be reductively explained all the way down to the physical (although it is admittedly a more arduous task than reductively explaining the simple rock). That is, externally observable or epistemically symmetric things all logically supervene the physical. The exclusive subjectivity and personal nature of experiential consciousness leads to epistemic asymmetry which is reason to believe that consciousness doesn't logically supervene the physical. Interestingly, there are intangible things that I would argue are not logically supervenient on the physical *because* they are dependent on consciousness, e.g. love, an experiential phenomenon.

I would now like to bring forth the infamous knowledge argument. The notable knowledge argument, first proposed by philosopher Frank Jackson, extends upon the idea of experiential epistemic asymmetry mentioned above and aids in the argument against the logical supervenience of consciousness on the physical. The knowledge argument states that Mary is a leading neuroscientist who is the foremost expert on color vision in the brain. She knows everything there is to know about color and how to perceive it, yet Mary has been raised her whole life in a grayscale room. She has never seen or perceived any other color. The purpose of this argument is to show that all physical facts that pertain to the perception of color don't account for the experience of *actually perceiving* a color. Even if Mary studied the physics of the color red, every context that red appears in, and all hues and shades of the color red, she has still yet to experience the color red. Suppose Mary is released from her grayscale room, to a world of all red. She finally gains that missing experience of red. Mary can now *imagine* red, even if she were to return to the grayscale room for the remainder of her life. In short, knowledge of conscious experience does not follow from knowledge of physical facts alone because physical facts alone don't account for experience.

In conclusion, consciousness does not logically supervene the physical due to its exclusive nature in the self. Phenomenal consciousness is impossible to analyze functionally, structurally, or quantitatively. The best one could do is describe how they feel as a conscious organism which doesn't truly define consciousness. My professor explaining how he consciously experiences does not allow me to consciously experience in the same way that he does. Yet, this is the best we're left with in understanding consciousness. As a result, the epistemic asymmetry of consciousness shows that the philosophical idea of solipsism is plausible, which in turn makes the concept of psychological zombies a conceivable logical possibility. The lack of a reductive explanation of consciousness in physical terms only helps to support my opinion: consciousness is *not* logically supervenient on the physical.

B. Opinion 2: Consciousness is logically supervenient on the physical

Nearly everything can be explained in physical terms, i.e. high-level phenomena can be explained via reductive explanation on a low-level set of physical properties. For example, the high-level phenomenon of the blue color of the sky is nothing more than a consequence of the physical characteristic of sunlight being scattered by the molecular gasses in Earth's atmosphere. Likewise, the green color of grass is nothing more than biologically understood chlorophyll that mostly reflects green light. The reductive explanations of the blue sky and grass can be continuously broken down without any extra "mapping," which is required in natural, or nomological, supervenience. Something is logically supervenient on the physical if there is a reductive explanation for it. The current lack of reductive explanation for phenomenal consciousness doesn't mean it doesn't exist as some might claim. In the future, consciousness may boil down to a complicated melting pot of biological properties, where the mapping laws of natural supervenience are not needed. We might learn that the physical entails consciousness in the same manner as the physical entails the general theory of relativity, and all that we are waiting for is a genius to make the discovery. As we believe, biological properties supervene logically on physical properties. Progress in technology could lead one to understand the exact biological structure that makes up each of a person's individual consciousness. However, an opponent might purport that just because you understand the arrangement of particles within a brain that creates consciousness, doubt still exists because you can't experience that instance of consciousness in the same way that the reader and myself may experience a rock. While this may be true, who says that scientific and technological progress may not lead to the idea of a human experiencing many consciousnesses at once or none at all?

Secondly, suppose a reductive explanation of consciousness exists that will never be understood by human beings, in a similar fashion to the way one knows a human being can't visualize the fourth dimension but knows it exists. Does the incapability of the human species' mental power prevent a true reductive explanation from being true, in a Platonist sense? Suppose an honest superbeing, vastly superior in thought than an average human, validated a reductive explanation of human consciousness to the physical? A proponent of the lack of logical supervenience of consciousness on the physical, Chalmers mentions the following within "A Conscious Mind": "it may be the case that some domains, such as those of sociology and economics, are so far removed from the simplicity of low-level processes that illuminating reductive explanation is impossible, even if the phenomena are logically supervenient. If so, then so be it: we can content ourselves with high-level explanations of those domains, while noting that logical supervenience implies that there is a reductive explanation in principle. Although perhaps one that only a superbeing could understand." I argue that just because a reductive explanation is out of reach doesn't eliminate the explanation from proving logical supervenience. In fact, the very terminology of reductive explanation leaves something to be desired. What are the rules of reduction? How would an explanation be validated? Is reductive explanation to be implicitly understood as anthropocentric? The cloud of ambiguity that shrouds the technique of reductive explanation may hurt this argument but would likewise hurt the argument for consciousness *not* being logically supervenient on the physical.

On a separate note, a rebuttal of the knowledge argument presented in the previous opinion will aid in the support of the idea that consciousness is in fact logically supervenient on the physical. The explanation provided by Jackson isn't quite satisfying enough to show that consciousness isn't logically supervenient on the physical. If Mary, the subject of the knowledge argument, truly knows everything there is to know about color, wouldn't her knowledge also encompass the functional feeling of the qualia that represents seeing a specific color? Mary would have a very low-level understanding of the brain's functionality, in which Mary would know the individual effects of viewing the color red. Mary would understand exactly how each neuron in her human brain and perhaps all human brains would register the experience of seeing the color red. On a low-level, Mary would understand the exact biological process that undergoes perceiving the color red. Mary would then be able to functionally understand the qualia without ever having to see the color red. Furthermore, when Mary is released from her grayscale room, one might argue she is not necessarily gaining a new fact of knowledge but rather learning the ability to imagine red. Consider a person blind from birth, and imagine that this person was immediately gifted with the ability to imagine the color red. Did this person gain a new fact of knowledge or a new ability? I would argue that the blind person learned a new ability in the same way that Mary learned the ability to imagine red. Mary holds restricted visual senses, much in the same way a blind person does, which is a result of the physical. To imagine the color red is nothing more than a biological configuration of the brain – something that Mary should admittedly know.

All in all, two separate approaches are used in my support of consciousness being logically supervenient on the physical. The first is understanding that reductive explanation is not constrained by the knowledge base that humans hold in the present or the mental capability of humanity. I conjecture that a plausible reductive explanation for consciousness on the physical may exist to the lack of our knowledge or understanding. Secondly, the knowledge argument allows us to pick apart the idea of qualia in support of physicalism, the idea that everything is either physical or logically supervenient on the physical. If to know everything about color includes knowledge of biological reaction that qualia performs on the brain for Mary to then visualize – say – the color red, then to learn qualia is nothing more than a physical fact as it is just a change to a biological structure. These arguments together comprise the opinion that consciousness is – in fact – logically supervenient on the physical.

C. Opinion 3: This conversation is irrelevant

We are presented with the following philosophical question: "*Is consciousness logically supervenient on the physical?"*. When you dissect this question, it truly is a fascinating exercise and can bring about confusion, thoughts, and feelings, as humans are conscious beings. I will initially address some notes on consciousness to bridge the gap between the posed question and scenarios that invalidate its discussion. The points in this opinion will join the discussion on consciousness to artificially intelligent systems to enforce a more practical context to the topic.

In regards to an artificially intelligent system, whether a system is conscious or not seems to be up for debate. Classically, my observation has been that people tend to discuss these matters from the point of view that our current technology is currently incapable of producing a conscious AI and that The Future will be the time when hard AI will produce a conscious being. "The Conscious Mind" by David Chalmers, gives excellent insight into the topic of consciousness, and his work will be referenced throughout this opinion. Language barriers are a big problem when discussing confusing matters, so Chalmers studiously hammers this topic "right off the bat".

Chalmers takes a page from Sutherland to define consciousness as follows, *"Consciousness: The having of perceptions, thoughts, and feelings; awareness. The term is impossible to define except in terms that are unintelligible without a grasp of what consciousness means. Many fall into the trap of confusing consciousness with self-consciousness to be conscious it is only necessary to be aware of the external world. Consciousness is a fascinating but elusive phenomenon: it is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written about it. (Sutherland 1989.)."* I believe that within this quote is where the true discussion needs to be had. Whether or not consciousness is supervenient on the physical is of no consequence to us as humans. Chalmers makes a point in "The Conscious Mind" that consciousness can be split into two categories: phenomenal and psychological. Reading further Chalmers does an excellent job of arguing that consciousness in fact does not supervene on the physical; however, the consciousness that is being explored is that of phenomenal consciousness.

This next portion may seem a bit crass or cold-hearted; however, I contend that when we debate the "consciousness" of an animal or AI we are really trying to understand if the entity in question has a sense of self. Does it have an Id, Ego, or Super-Ego? Can it distinguish itself from the external world? Chalmers defines self-consciousness as follows, "*This refers to our ability to think about ourselves, our awareness of our existence as individuals and of our distinctness from others. My self-consciousness might be analyzed in terms of my access to a self-model, or my possession of a certain sort of representation that is associated in some way with myself. While it is plausible that some degree of phenomenal consciousness is possessed by animals much less sophisticated than ourselves, it may well be that self-consciousness is limited to humans and a few species of animals."* This is the true conversation and for it, personal morality will be shed to attempt to address concepts in an objective manner. Some may wonder why the topic of consciousness in terms of phenomenal consciousness can be cast aside so easily. The answer is simple and can be explored through two different avenues.

Suppose a situation that is familiar to dog owners but can be extended to other animals. Imagine the following in a third-person point of view. An individual comes home from work and their dog greets them at the door. The dog is displaying a positive reaction to the owner's presence. Upon noticing the dog has not made a mess of their home, the owner responds positively to the dogs and gives them a treat. The dog displays increased excitement from receiving the treat and curls up next to the owner after devouring the treat. This scenario is in direct contrast to how the dog would react if the owner had come home, found a mess, and disciplined their dog. The dog has a phenomenally conscious experience in either case (positive or negative) that reinforces similar behaviors. If this scenario is possible for a dog, it is not a large jump to overlay this observation on livestock such as cattle or swine. Indeed, I have personal experience with both and would argue that they undergo these experiences from input stimuli. From this scenario, it is likely to conclude that these class of animals have a conscious experience, yet as humans, we routinely slaughter them (livestock) for consumption with little or no thought to the contrary.

Another scenario that should be considered rings closer to home than animals. Think back to your very first memory and try to picture it in your mind. Think of the sensations involved with it and then realize that from the first-person point of view, your sense of self began with that day. There is no debate that as a human being, you are a conscious being that processes both phenomenal and psychological experiences. Often an entire year of our lives pass without any recollection as to the events, pleasures, and pains, yet we are considered conscious. From a third-person point of view and before an infant has their first memory and establishes a sense of self, they cannot be considered empirically different from that of a “zombie” or an animal. Logically, we can establish that they react to a stimulus that they perceive from sensory input but that is all. At that point, they can be simply viewed as a processor of input data and state that can be “turned off” with little objection from the subject, despite their conscious state. Therefore, the debate of phenomenal and psychological consciousness “holds no water” in regard to AI.

Objectively the ending of a human life, at any stage of development, is considered morally reprehensible in human society because even in their infancy, humans know from their own experience that they have the potential for self-consciousness and autonomous interaction with the external world. Some may argue that this is an argument for when life begins or what it means to exist. That is a valid point; however, it solidifies the notion that the conscious state of a physical object is not the question that needs debate. Taken to the extreme and setting aside politics, abortion is the termination of a human fetus in utero before it can survive outside of its host environment. Despite knowing the potential of the developing human and the fact that neurological processes necessary for a consciously phenomenal experience exist, humans regularly decide that terminating the developing human is justified since it is not known scientifically when life “begins”.

In summary, the question of consciousness logically supervening on the physical is a waste of time when considering the true ethics of creating a self-aware AI that has created its own identity. Humans have routinely shown that they are willing to end life or the potential for life of a being that is guaranteed to have a higher level of consciousness than that of a developmental AI. Since this is the case, why would humans care about the conscious state of less developed intelligence? The movie *Ex Machina* accurately depicts the point of view that is trying to be stated. *Ex Machina*’s plot defines a protagonist that empathizes with a self-aware AI despite the physical differences and the unknown of the AI’s capacity of empathizing with the protagonist. Additionally, the creator of the AI is able to disengage the simulated human behavior of the AI from the AI’s physical machine in order to overcome the ethics proposed by termination of the AI. This opinion takes the stance that humans and society will always hold the point of view of the creator based on current human-to-human inhumanity.

IV. Relevance of Logical Supervenience of Consciousness and AI

It's not a surprise that as artificial intelligence technically progresses into the future, nearly all other realms of life will be affected. Apart from the vast automation of current-day jobs, which will cause a considerable shift in required skillsets for employment or the changes in social stratification that may occur, the laws around the world will have to consider how to treat artificial intelligence as the differences between AI and humans grow thinner and thinner. The discussion about consciousness and whether it logically supervenes on the physical may seem needlessly pedantic or "debate for the sake of debate", but the way in which we define consciousness can very well influence the way we define governmental rights.

Placing consciousness "over and above" the physical via natural supervenience assigns a label of intrinsic worth. Beings with phenomenal consciousness no longer can be partitioned into the same class as a cut of beef or a classical microprocessor which we currently believe supervene upon the physical, due to reductive explanation. This implicit worth can be used within a legislative framework regarding human and non-human rights. If a conclusion can be reached that this machine has no phenomenal conscience, then that particular machine may not be given rights like a human. Alternatively, imagine that we find that a machine *does* employ some degree of phenomenal conscience! Our world view of AI would change – perhaps even along with the world view of our own species. For, once we consider the way in which we treat machines may we begin to understand the poor way that we've been treating each other. The continuous debate regarding the modal force behind the supervenience of consciousness may give way to transformative discoveries or thought lessons regarding how to treat one another, animals, machines, and – for all anyone knows – even a rock, because all anyone knows about consciousness is *very* little.

V. Data Analysis & Feature Selection

A. Progressive Rock Characteristics

The music genre of progressive rock, known as prog-rock, blends together jazz, classical music, and rock in a virtuosic fashion. Typically, songs that belong to the progressive rock genre are longer in nature, feature numerous unique song segments, and are arranged in unorthodox manners to eschew common verse/chorus formats. Prog-rock songs explore changes in meter and contain large variations in key or modality compared to most other music genres. The instrumentation of prog rock music often spanned a wider array of acoustic and electronic instruments as compared to traditional rock. In addition, lyrical content (or lack thereof), album artwork, and the focus on concept albums over singles are distinguishing features that separate prog-rock from other music genres.

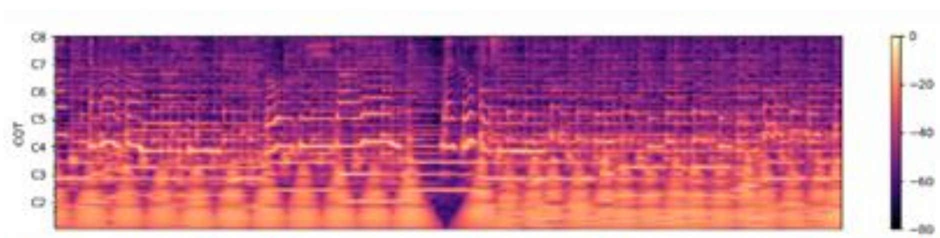

Fig. 1. Visual depiction of an audio file after it has gone through a Short-Time Fourier Transform, often referred to as the STFT of a signal.

B. Feature Extraction, Analysis, and Selection

Our feature selection approach entailed finding all extractable, repeatably calculable, and potentially helpful features extant in raw audio data, namely MP3 files. *Fig. 1* shows a visual depiction of the Short-Time Fourier Transform (STFT) of an audio signal as a spectrogram. Without going down the road of digital signal processing, the STFT performs the conversion from the time domain into the frequency domain of the original signal over a finite amount of time. STFT contrasts with the traditional Fourier Transform that normally computes the entire input signal into the frequency domain. The frequency domain is responsible for a wide variety of spectral features used in music information retrieval, e.g. spectral contrast and roll-off. The LibROSA Python music information retrieval (MIR) package, which implements STFT and FT algorithms, was utilized to process song files and extract several of the features found in *Table 1*. *Table 1* shows the list of features (and their corresponding descriptions) that were extracted and tested over the course of the development cycle. Through tests, analysis, and trial and error, we found the most meaningful features out of the array listed in *Table 1*.

Feature	Description
Chroma Constant Q Transform (CQT)	Translate time-domain signals to the frequency domain; very similar to the Fourier Transform
Chroma STFT	A chromagram, computed using STFT, which shows the importance of the twelve pitch class profiles of a given waveform
Mel-Frequency Cepstral Coefficients (MFCC)	Set of features that are used to describe the timbre and shape of a spectrum
Polynomial	An *n*-th order polynomial that is fitted to columns of a spectrogram
Roll-Off	The frequency at which a majority of the spectral energy is under; a measure of the shape of a waveform
Root-mean-square energy (RMSE)	Effectively measures the area under the waveform which represents the power of amplification of a musical signal
Spectral Bandwidth	The wavelength interval over which the magnitude of spectral components is equal to or greater than a fraction of the maximum spectral component
Spectral Centroid	Represents the “center of mass” of a spectrum of sound that is used as a measure of brightness and timbre
Spectral Contrast	The frequency difference between spectral peak and contrast
Spectral Flatness	A decibel-measure of noise within an audio spectrum
Tempo (Average)	Average beats-per-minute of a segment of a song
Tempo (Variance)	A measure of the frequency of the change in meter of a song segment
Tonnetz	The tonal centroid features of the tonnetz, a lattice-like structure used represent tonal space
Zero Crossing Rate	The frequency at which a signal switches from positive to negative; excellent measure of frequency of percussive sounds

Table 1. Music Information Retrieval (MIR) Features and Descriptions

Given an audio file, each of the above features are computationally feasible and can be used to describe different portions of the signal. Initially, we extracted all features using LibROSA. After multiple analyses were performed on tests using the training data with this expansive feature set, two distinct conclusions became apparent. First and foremost, even with splitting the songs into “separate” songs to increase the amount of training samples, the training data is inherently small compared to the possible (practically infinite) test inputs. The training dataset contains an extreme class imbalance, where 73 songs were considered prog and the remaining 302 songs were of various music genres altogether labelled non-prog. This imbalance in the training examples obviously creates a bias in the classifier towards the non-prog case. Secondly, some features proved insufficient in providing discriminatory information.

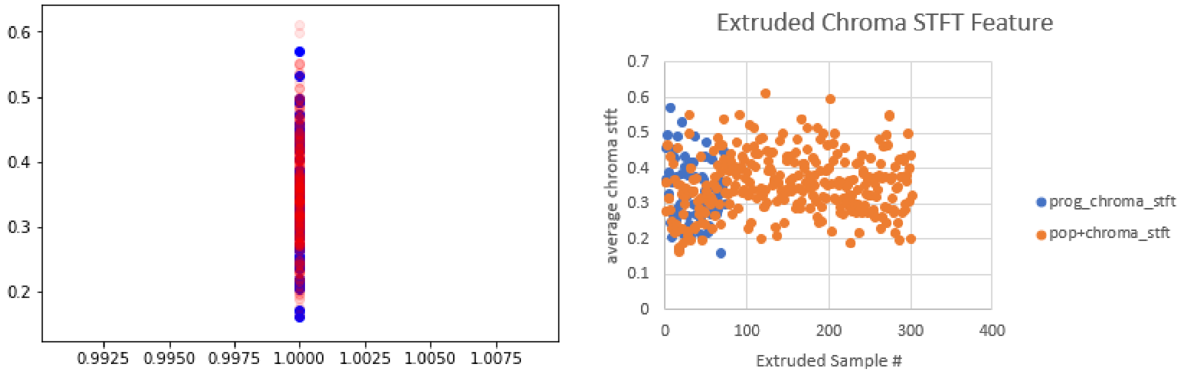

Fig 2. Example of the Chroma STFT Feature for Illustration. Left: 1-dimensional representation of the Chroma STFT feature. Blue dots represent the prog songs, and red dots represent all other music genres. Note that the red dots were so dense that the alpha value had to be set to "0.1" so that the prog-song datapoints could be made visible. Right: Chroma STFT data extruded along a fake sample axis. The y-axis data is identical to the plot on the left.

It is important to note that *Fig. 2* is being displayed as an extruded version of a 1D feature set. Specifically, *Fig. 2* shows Chroma STFT feature which focuses on the harmonic segment of an audio file. When the data is compressed down to its 1D representation, the prog/non-prog data becomes indistinguishable and therefore will not significantly aid in the discrimination of the target music genre.

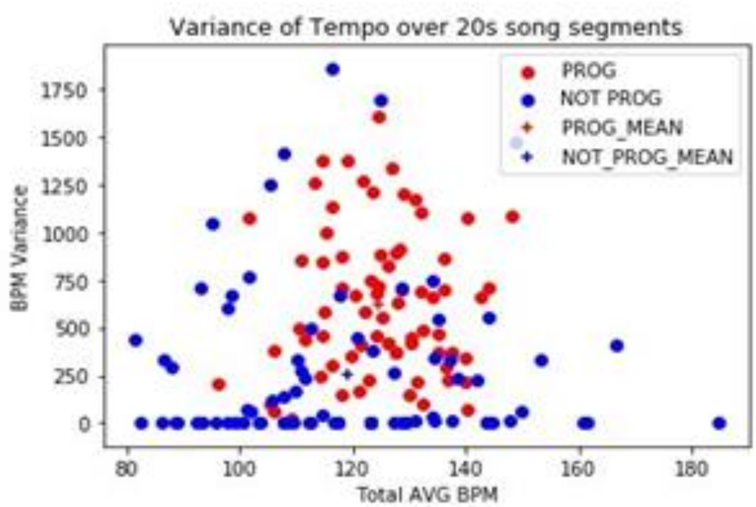

Fig 3. BPM Variance vs. Total Average BPM. Example of a feature that contains useful information for discrimination. Here the blue and red '+' denote the means of all songs of a particular class.

On the other hand, *Fig. 3* shows a meaningful separation between the classes of prog and non-prog. A number of non-prog songs display little tempo variance over the course of the track, when compared to prog songs. Through similar evaluation of individual features as well as trial and error, it was chosen to move forward with using the MFCCs over a time series of a single song as the total feature set. Using features focused on time-series complimented the success of certain classification models, namely the long short-term memory (LSTM) model, which is detailed in Section VII.

C. Discussion of Mel-Frequency Cepstral Coefficients (MFCC)

Now we will present a brief introduction to MFCCs since we chose to highlight them as the basis of our feature set. The initial interest in MFCCs arose from research on audio classification and their heavy use in speech recognition systems. The MFCCs physically represent the amplitudes of peaks of a spectrum-of-a-spectrum after the STFT has been mapped to the range of the Mel Scale.

$$m = 2595\ log_{10}\left(1 + \frac{f}{700}\right) \tag{1}$$

Eq. 1 above shows the conversion formula to convert from a given frequency to "mels," the unit of measure of the Mel Scale. This is not the only version of the formula available, although the only difference between formulae comes down to the coefficient preceding the *log* function. This scale is particularly interesting because it maps frequencies into the audible range of human hearing, typically 20Hz to 20 KHz. Once the MFCCs have been found, they can be used to model a particular sound. In a sense, MFCCs map an aural fingerprint to a sound, which explains their effectiveness in speech recognition applications. It can also be said that MFCCs succinctly encode the timbre, or the unique texture, of a sound. As one might expect, unique mathematical coefficients that describe the timbre of similar sounds are excellent features to utilize in a musical genre classification model.

D. MFCC Feature Space

We initially considered just the mean of the MFCC components for training our initial models and soon realized the importance of capturing the whole time series data. Next, we generated time series feature vectors by specifying a hop length parameter. The hop length parameter specifies the duration of samples from a specific song segment. Then, LibROSA generates the features at the specified hop length for each song. In order to have uniformity in size among features of all the songs, the hop length was restricted to the minimum song length within the whole training set. As a result, performance hit because time series data of longer songs (especially longer prog songs) were not considered. We used this input to train models based on Random forests, CNN and LSTM.

VI. Classification Models Summary & Performance on Training & Testing Sets

It should be noted that we took a competitive approach to developing classifiers in our group, and multiple methods were employed in order to determine the best classifier amongst us. The collection of approaches below define the development that culminated in our final, best-performing, model.

A. Fully Connected Neural Network

Our first approach was to take the mean of the features from *Table 1* for each song. The result is a 1D vector space for each song. Upon training a standard neural network with 4 fully-connected dense layers, the model achieved close to 100% classification efficiency in training data, after about 50 epochs. However, the model was not classifying the validation data to the expected level. Thus, we concluded that we shift our focus to time series vectors instead of relying on scalar mean values.

We realized that the specific input feature set needed refinement. The refinement of the feature set was performed via empirical testing on the initial NN model since it was faster to train. The testing was done by individually eliminating specific features from the input set and comparing the output to the nominal case (all features included). The empirical testing led to a conclusion that most of the classification information was developed from the MFCCs; therefore, the MFCCs remained as the only features in the time-series feature vectors that were used in the final model. Specifically, the top thirteen MFCCs were selected as meaningful features that affected the classification accuracy the most, when the songs were divided by a hop length of 1024 samples.

	Predicted: Non-prog	**Predicted: Prog**
Actual: Non-prog	52	6
Actual: Prog	16	1

Table 1. Confusion Matrix on Validation Set

Code Link:

- https://uflorida-my.sharepoint.com/:u:/g/personal/s_gothi_ufl_edu/Eay6X0wbD2lPnOIgD46c3h8BTFV6lcArFNJHQchpJ0nW2A?e=xA2xTo

B. Convolutional Neural Networks (CNN)

We tried to execute a CNN model with the above time series features using a series of two convolutional layers followed by a couple of dense layers. Actions were taken to mitigate the issue at hand, like adding dropout layers to decrease the degree of overfitting taking place. Additionally, we experimented with batch normalization layers to no avail. The result ended up in the similar manner as the random forests: every example was being labelled the same.

Another technique that was employed was to calculate the spectrogram of song segments as seen in *Fig 1*, save them as images, and execute a CNN with two convolutional layers on the resultant spectrograms; however, this visual approach to sound classification did not provide acceptable performance. The pattern of uniform non-prog classification continued with this model too!

Code Link:

- https://uflorida-my.sharepoint.com/:u:/g/personal/s_gothi_ufl_edu/ETaJorq1jRVCrJQIECC4bJsBDUKoRac_H8QYeiXvG7dKqQ?e=Dtxlwn

C. Random Forests (RF)

The possibility of using a Random Forest (RF) model was attractive for two main reasons. Primarily, RF models are very simple to develop, train, and test. In addition, RF models, if used correctly, are theorized not to overfit the data. Therefore, RF models can establish a viable benchmark for other classification models to follow. RF models were separately trained on standard input feature vectors and time-series feature vectors to find a baseline performance accuracy. *Fig 4* below shows example outputs of the 33D standard average features for a single song along with the description of the model. *Fig. 4* and *Fig 5* illustrate that despite reaching 79% and 77% accuracy, respectively, on the validation set, it is mostly meaningless because the validation set consisted mostly of non-prog songs. Due to this, the RF model confirmed our initial findings that the input features are not providing enough discriminable information as mean scalar values.

```
parameters = {'bootstrap': True,
              'min_samples_leaf': 3,
              'n_estimators': 5000,
              'min_samples_split': 2,
              'max_features': 'sqrt',
              'max_depth': 50,
              'max_leaf_nodes': None}
```

```
Accuracy on 90/10 train split:  0.92105263157894
[[34  0]
 [ 3  1]]
Score on the Validation Set =  0.793478260869565
[[71  0]
 [19  2]]
```

Fig 4. Non-Time Series RF Model Parameters & Training and Validation Confusion Matrices

A separate RF model was executed using the time series data with an input feature shape of (375, 1590, 33). The performance of the model is shown in *Fig. 5* below.

```
parameters = {'bootstrap': True,
              'min_samples_leaf': 3,
              'n_estimators': 2000,
              'min_samples_split': 10,
              'max_features': 'sqrt',
              'max_depth': 50,
              'max_leaf_nodes': None}
```

```
Accuracy on 90/10 train split: 0.8947368421052632
[[34  0]
 [ 4  0]]
Score on the Validation Set: 0.7717391304347826
[[71  0]
 [21  0]]
```

Fig 5. Time Series RF Model Parameters & Training and Validation Confusion Matrices

Code link:

- https://uflorida-my.sharepoint.com/:u:/g/personal/s_gothi_ufl_edu/EZIOYZBlN6pKsEOryNj_2PcBHyXTkWGPULZC6lzSBbHrFA?e=uq6TLa
- https://uflorida-my.sharepoint.com/:u:/g/personal/s_gothi_ufl_edu/ESSJR4H3rhVHoquZkW_AfLYBBdhjL7wRxVBK2rCNMwn01Q?e=I7UUqu

D. Long Short-Term Memory (LSTM)

The poor performance from previous models and the nudge from Professor Rangarajan led us to pursue LSTM. Research shows that LSTM works effectively with time series data, thus we wished to explore its merits as a music genre classifier. Upon prepending an initial LSTM that fed into the NN, the classification accuracy still had not improved, and the confusion matrix indicated bias towards the non-prog class. Given this new information we decided to further refine our initial input features.

LSTM needs 3-D input matrix, where the 3rd dimension is the time series samples. Since song lengths vary, we had to try different approaches.

Approach 1:

First, we extracted the time samples for each song and tried to identify the song with the minimum number of time samples and extracted that many time samples from each song. This approach neglected some major parts of some songs, since the lowest length is around 1 minute and prog songs can run *very* long.

	Predicted: Non-prog	**Predicted: Prog**
Actual: Non-prog	71	0
Actual: Prog	21	0

Table 2. Confusion Matrix on Validation Set

Code Link:

- https://uflorida-my.sharepoint.com/:u:/g/personal/s_gothi_ufl_edu/EdPUmQEilUZJqU9mDH8gnS4B7nYhLZCGrBfZOKuNbT0ilQ?e=oORq81

Approach 2:

A song was divided into a standardized hop length, and each feature was extracted for that particular audio segment where each audio segment was given a bin identification number out of 200. Each feature was then calculated from the bins in order to produce the time series features of a song. We then ran multiple tests where the hop-length was varied and picked the hop-length that delivered the best efficiency. A hop length of 1024 yielded the best efficiency.

With the top MFCC components and the buckets of features for each song, we ran with several iterations of multiple layers of LSTM. Observations showed that having two layers gave us the optimal classification accuracy on training and validation data. With more than two layers, the model was overfitting towards labelling everything as prog. Additionally, a bidirectional layer was implemented into the model. By including the bidirectional layer, it's as if the model is training not one, but two LSTM's where one LSTM is training on the original input and the second on the reversed input! Although it's safe to say most humans think of music in the sense of beginning-to-end, a machine may gain more insight into classification by also viewing audio from end-to-beginning as well! The extra context of the bidirectional layer increased the success of the LSTM model.

	Predicted: Non-prog	**Predicted: Prog**
Actual: Non-prog	54	17
Actual: Prog	3	18

Table 3. Confusion Matrix on Validation Set

	Predicted: Non-prog	**Predicted: Prog**
Actual: Non-prog	88	12
Actual: Prog	38	65

Table 4. Confusion Matrix on Test Set

Code Link:

- https://uflorida-my.sharepoint.com/:u:/g/personal/s_gothi_ufl_edu/Ec-_M2slQUhFjS_D9MPpwKkBaUtRVyFX--a9xsjY2KUoGw?e=uHfCXZ
- https://uflorida-my.sharepoint.com/:u:/g/personal/s_gothi_ufl_edu/EeUgNYrYPSlGiAUi8D7nlNsBkbuBOp5MTSpCH6eMmFKboA?e=2Uod7H

Approach 3:

The 200 samples felt low, so we combined multiple bins into one by taking means of consecutive samples. This way our bin size remains the same while avoiding the vanishing gradient problem. At the same time, we were able to create a bin representing larger segment of song.

	Predicted: Non-prog	**Predicted: Prog**
Actual: Non-prog	62	9

Actual: Prog	14	6

Table 5. Confusion Matrix on Validation Set

Code link:

- https://uflorida-my.sharepoint.com/:u:/g/personal/s_gothi_ufl_edu/EVLvdwJqvY5JuzANe0q1V-QBwy85ytFYZz46-thraXjjbQ?e=wO7O5L

VII. Conclusion

Interestingly, the solution to musical genre classification shares a significant overlap with the solution speech recognition. The approaches to both music genre classification and speech recognition seem to favor time series input vectors, LSTM models, and bidirectional layers. Because LSTM is so well-suited for time series data, it expectedly performed the best out of all the attempted approaches, finally overcoming some issues that we had been facing. Our group faced the reoccurring obstacle of models completely overfitting/underfitting and outputting one label in a "one size fits all" approach to classifying input songs. The imbalance of prog songs to non-prog songs certainly factored into the source of this issue. Although we would have hoped for a more balanced training set, we understand that in the real-world, the provided training data that we are to work with will never be totally ideal. It's in these scenarios where we must make do with what we have, using our prerequisite knowledge to carve out a best path and overcome any issues.

Out of all the models we have trained and tested, we felt that LSTM based Approach 2 gave the best result on both validation and test set. Final testing accuracy is 76.5% and the confusion matrix is depicted in *Table 4*

To view all project code, visit the following link:
https://uflorida-my.sharepoint.com/:f:/g/personal/s_gothi_ufl_edu/EkLuQIeJF3lGnguokgs_BsIBYjsLYEKysM3PKyL0GHyCxw?e=YVYMSP

VIII. References

[1] Chalmers, D. (2007). The conscious mind. New York: Oxford University Press.
[2] "Characteristics of progressive rock", *En.wikipedia.org*, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Characteristics_of_progressive_rock. [Accessed: 03- May- 2019].
[3] F. Jackson, "Epiphenomenal Qualia", *The Philosophical Quarterly*, vol. 32, no. 127, p. 127, 1982. Available: 10.2307/2960077.
[4] Librosa.github.io. (2019). LibROSA — librosa 0.6.3 documentation. [online] Available at: https://librosa.github.io/librosa/ [Accessed 2 May 2019].
[5] Plato.stanford.edu. (2019). Supervenience (Stanford Encyclopedia of Philosophy). [online] Available at: https://plato.stanford.edu [Accessed 2 May 2019].
[6] Practicalcryptography.com. (2019). Practical Cryptography. [online] Available at: http://practicalcryptography.com/miscellaneous/machine-learning /guide-mel-frequency-cepstral-coefficients-mfccs/ [Accessed 2 May 2019].
[7] How to Prepare Univariate Time Series Data for Long Short-Term Memory Networks [online] Available at: https://machinelearningmastery.com/prepare-univariate-time-series-data-long-short-term-memory-networks/ [Accessed 28th April].
[8] https://towardsdatascience.com/whats-wrong-with-spectrograms-and-cnns-for-audio-processing-311377d7ccd