Nonlinear Brain Tumor Model Estimation with Long Short-Term Memory Neural Networks

Jiashu Guo, Zhengzhong Liang, Gregory Ditzler
Dept. of Electrical and
Computer Engineering
The University of Arizona
Tucson, AZ, USA

Nidhal C. Bouaynaya
Dept. of Electrical and
Computer Engineering
Rowan University
Glassboro, NJ, USA

Elizabeth Scribner
and Hassan M. Fathallah-Shaykh
Department of Neurology
University of Alabama at Birmingham
Birmingham, AL, USA

***Abstract*—Gliomas are malignant brain tumors that are associated with high neurological morbidity and poor outcomes. Patients diagnosed with low-grade gliomas are typically followed by a sequence of measurements of the tumor size. Here, we show the promise of Long Short-Term Memory Neural Networks (LSTMs) to address two important clinical questions in low-grade gliomas: 1) classification and prediction of future behavior; and 2) early detection of dedifferentiation to a higher grade or more aggressive growth. We use a system of partial differential equations (PDEs), from our earlier work, to generate simulated growth of low-grade gliomas with different clinical parameters. We design an LSTM network to solve the inverse problem of PDE parameter estimation. We find that accuracy increases as a function of the number of tumor measurements and perplexity can also be used to detect a change in tumor grade. These findings highlight the potential usefulness of LSTMs in solving inverse clinical problems.**

I. Introduction & Relationship to the State-of-the-Art

Gliomas are malignant brain tumors that continue to be associated with significant neurological morbidity and poor outcome. The World Health Organization (WHO) classifies adult gliomas into low-grade (grade 2), anaplastic (grade 3), and glioblastoma multiforme (GBM, grade 4). Each group is further characterized by pathological and radiological features [1]. Brain invasion (motility), Angiogenesis (new blood vessel formation) and proliferation are characteristic features of gliomas [1]–[4]. Brain invasion by glioma cells leads to neurological morbidity. Glioma motility is believed to be mediated by two mechanisms. The first, concentration-driven motility (dispersion), drives invasive cells from high to low concentrations; its rate being variable between tumors [2], [3]. The second mechanism, hypoxia-driven motility, is associated with increased motility under hypoxic conditions [4]. We recently reported a nonlinear mathematical model of glioblastoma multiforme growth, angiogenesis, and invasion at the scale of clinical magnetic resonance imaging (MRI) that includes both concentration-driven and hypoxia-driven motility [5], [6]. The type and rate of motility are associated with distinct progression patterns and survival times when the patients are treated with bevacizumab [5]. Furthermore, computational trials have replicated the overall survival times of GBM patients treated with bevacizumab or by Tumor Treating Fields [6]. The mathematical model also replicates the biological behavior, radiological and pathological features of grades 2-4 gliomas and transition to a higher grade [7].

Clinicians caring for patients with low-grade gliomas are confronted with two important problems; the *first* is the future growth potential of a particular tumor. The *second* is the early detection of a change in tumor grade that leads to more aggressive growth. Typically, these patients are followed by serial magnetic resonance imaging (MRI) that permits the measurement tumor mass. A novel methodology is needed to help classify the long-term behavior of these tumors and uncover a change in grade. Such advancements will improve patient outcomes by optimizing the timing of therapy.

The mathematical model of gliomas is a coupled system that includes a partial differential equation (PDE) [5]. Translation of the model to the care of patients diagnosed with gliomas, requires solving the inverse problem of parameter estimation given a time-series set of measurements from clinical tumors. Classical methods for parameter estimation of PDEs include proper orthogonal decomposition (POD). POD was initially presented by Kari Karhunen and Michel Loéve independently of one another in the 1940s, and therefore, it is also known as the Karhunen-Loéve Expansion (KLE) [8]. Lawrence Sirovich later developed the method of "snapshots," or observations, to the KLE, which reduces the number of eigenvectors by excluding eigenvectors with eigenvalues less than a certain value [9]–[11]. The goal of the POD method is to produce a reduced-order model of a system, which can then be used to solve the inverse problem of parameter estimation in a set of differential equations [12]–[16]. An important limitation of the POD method is that it is computationally intensive.

Artificial neural network modeling has been applied in many areas, including signal processing, pattern recognition, system identification and control [17]. Although the study of recurrent neural networks (RNNs) for parameter estimation in dynamical systems, notably state-space models, was established before the hype of deep learning [18], [19], its application to large clinical datasets is very recent [20], [21]. To our knowledge, there is no literature on estimating parameters of PDE models using recurrent neural networks nor their applications to brain tumor modeling. In this contribution, we

Corresponding author: ditzler@email.arizona.edu

propose a Long Short-Term Memory (LSTM) recurrent neural network as a novel solution to the inverse problem associated with modeling gliomas. Our solution has the potential of classifying low-grade gliomas by estimating their rates of growth, migration, and angiogenesis. Furthermore, through simulations, we find that perplexity can be applied to detect a change in tumor growth potential.

II. Modeling Tumor Growth

A. Model Description

The model and numerical methods are detailed in our recent publication [5]. Briefly, several cellular types are considered; proliferative cells, denoted by P, are glioma cells that are actively dividing and do not move. Invasive cells, I, migrate but do not divide. Brain matter cells, B, begin at an initial fixed concentration; they do not divide or migrate. If the hypoxia is severe, cancer cells P and I, as well as brain cells B, eventually die; this is called necrosis N. The total number of cells, C, is taken as the sum of P, I, B, and N (see Equation (5)).

Hypoxia causes P cells to stop dividing and switch to I cells, which can leave the core of the tumor and invade the brain. After traveling in the brain and reaching a favorable local environment, I cells stop their movement and become proliferative again, leading to tumor growth. These assumptions are called the "go-or-grow" phenotype [7]. We allow the available oxygen/nutrient supply to vary indirectly with the total concentration of cells C (Equations (5)-(6)); furthermore, angiogenesis elevates a key local hypoxia threshold, which varies directly with P (Equation (9)).

The local hypoxic state H (Equation (6)) in any specific brain location is a function of C, and we identify two main critical thresholds: C_{hyp} (hyp for hypoxia), when cells begin slowing their growth rate and switching from one phenotype to another, and C_{ltm}, when cells begin to die. The mitotic rate M varies spatially depending on C (*i.e.*, hypoxia, Equation (7)). P cells divide at their maximal rate when $C < C_{hyp}$. The mitotic rate decreases and is inversely proportional to the local hypoxic state when $C_{hyp} \leq C < C_{ltm}$; and it eventually vanishes when C exceeds C_{ltm}, which triggers necrosis (γF, Equation (8)). Angiogenesis elevates the thresholds C_{hyp} and C_{ltm} as a function of P (see Equations (9)-(10)), thus supporting a denser tumor before the death rate reaches its maximum γ.

The concentration of P cells is modeled by Equation (1), which corresponds to our assumptions: P cells divide, switch to an invasive phenotype ($P \rightarrow I$) or die as a function of the level of local hypoxia (γFP). Furthermore, they appear and subsequently divide when invasive cells switch back to the proliferative phenotype ($I \rightarrow P$), *ie* when I cells reach new regions of low cellular density (and hence higher levels of nutrients). When the hypoxic state is high, *i.e.*, H is close to 1, the transition of P to I ($P \rightarrow I$) occurs at the maximum rate α. Conversely, $I \rightarrow P$ occurs at a rate $\beta(1-H)$, which is elevated when hypoxia is low, $(1-H) \approx 1$, and depressed when hypoxia is high, $(1-H) \approx 0$.

I cells are only produced by proliferative cells under hypoxic conditions. The first migration term, $\delta \nabla \cdot (D \nabla I)$, describes passive diffusion (PD) or concentration-driven motility. Another parameter, D, varies spatially to replicate the increased rate of movement of cancer cells along white matter tracks in the brain. The second migration term, $\eta \nabla \cdot (I \nabla B)$, reflects the preferential movement of cells in the direction towards areas with the highest number of healthy brain cells (Active Transport, AT); the speed of migration is proportional to the concentration of I cells and the gradient of B, ∇B.

B. Model Equations

Proliferative Cells:

$$\partial_t P = \underbrace{MP}_{\substack{\text{Net production} \\ \text{of P cells}}} - \underbrace{\alpha HP}_{\substack{\text{Conversion of} \\ \text{P cells to I} \\ \text{during hypoxia}}} + \underbrace{\beta(1-H)I}_{\substack{\text{Conversion of} \\ \text{I cells to P} \\ \text{during normoxia}}} - \underbrace{\gamma FP}_{\substack{\text{Necrosis of} \\ \text{P cells} \\ \text{during hypoxia}}} \tag{1}$$

Invasive Cells:

$$\partial_t I = \underbrace{\delta_I \nabla \cdot (D \nabla I)}_{\substack{\text{Passive diffusion} \\ \text{of I cells}}} - \underbrace{\eta \nabla \cdot (I \nabla B)}_{\substack{\text{Active transport} \\ \text{of I cells}}} + \underbrace{\alpha HP}_{\substack{\text{Conversion of} \\ \text{P cells to I} \\ \text{during hypoxia}}} - \underbrace{\beta(1-H)I}_{\substack{\text{Conversion of} \\ \text{I cells to P} \\ \text{during normoxia}}} - \underbrace{\gamma FI}_{\substack{\text{Necrosis of} \\ \text{I cells} \\ \text{during hypoxia}}} \tag{2}$$

Brain Cells:

$$\partial_t B = \underbrace{-\gamma FB}_{\substack{\text{Necrosis of} \\ \text{brain cells} \\ \text{during hypoxia}}} \tag{3}$$

Necrotic Cells:

$$\partial_t N = \underbrace{\gamma F(B+I+P)}_{\substack{\text{Conversion of P, I, and B} \\ \text{to necrotic cells} \\ \text{during hypoxia}}} \tag{4}$$

Total cell concentration:

$$C = P + I + B + N \tag{5}$$

Measure of Local Hypoxia:

$$H = \begin{cases} 0, \text{ if } C < C_{hyp}. \\ 1 - \frac{C_{ltm} - C}{C_{ltm} - C_{hyp}}, \text{ if } C_{hyp} \leq C \leq C_{ltm}. \\ 1, \text{ if } C \geq C_{\text{ltm}}. \end{cases} \tag{6}$$

Mitotic Rate:

$$M(H) = \tau(1-H) \tag{7}$$

Rate of necrosis:

$$\gamma F(C) = \gamma \frac{1 - \tanh(30(C_{ltm} - C))}{2} \tag{8}$$

Hypoxic threshold:

$$C_{hyp} = \sigma[\log(1 + P)] + \Omega, \tag{9}$$

where $\sigma = 1.5$ to simulate angiogenesis, and $\sigma = 0$ to simulate anti-angiogenesis.

Necrotic threshold:

$$C_{ltm} = C_{hyp} + \Phi \tag{10}$$

A description of the parameters is as follows: α and β represent the transition rates from $P \rightarrow I$ and $I \rightarrow P$, respectively. τ is the mitotic rate of P cells. δ is the diffusion coefficient of I cells. D is the diffusion tensor in the brain. γ is the rate of necrosis. η represents the rate of active transport (hypoxia-driven motility) of I cells. Ω is the initial hypoxic threshold. σ represents the rate of angiogenesis. Φ is the fixed difference $C_{ltm} - C_{hyp}$.

C. The Inverse Problem and Parameter Estimation

We have *two objectives*; the *first objective* is to estimate the parameters of the system of equations when it models grade 2 gliomas; simulations of tumor mass were generated by titrating the rates of proliferation (τ), concentration-driven motility (δ), and angiogenesis (σ), as shown in Table II, such that they are associated with the pathological and radiological features and the survival times of grade 2 gliomas. We assume that $\eta = 0$ since hypoxia-driven motility (active transport) is a hallmark of higher grade tumors, (*i.e.*, GBM). This analysis sets the stage for future work on parameter estimation using clinical data. Nonetheless, an important and fundamental question arises: how many measurements are needed to obtain a high degree of confidence that the parameter estimation is accurate?

The *second objective* is to derive a method that detects a change in model parameters. The rationale is based on the fact that some grade 2 gliomas dedifferentiate to a higher and more malignant grade. This progression is typically associated with a change in the rates of proliferation, motility or angiogenesis (*i.e.*, parameters). The application of this objective is earlier detection of a change in grade and therefore timely therapeutic intervention.

III. Implementation and Results

In this section, we discuss how the data are collected, pre-processed and learned by an LSTM recurrent neural network for the parameterization problem presented in Section II.

A. LSTM Implementation

In this section, we briefly review the LSTM model and how it relates to the task of modeling tumor growth and parameterizing the dynamical system. Our objective is to solve the inverse problem that will identify the parameter of the PDE given a sequential set of measurements of a patients brain tumor size. This inverse problem can be solved using an LSTM recurrent neural network, which has been shown to be good at capturing long-term dependencies between historical input and current input data [22], [23]. The basic unit of LSTM network is the memory block as shown in Figure 2 as opposed to a standard neuron that is used in an RNN and the memory cell containing the information over a longer period than a standard neuron. The input to an LSTM block is the signal coming from the input sequence $\mathbf{X}_t$ and the recurrent feedback from the hidden layer $\mathbf{h}_{t-1}$, which are added together after a projection with matrices $\mathbf{W}_z$ and $\mathbf{H}_z$, respectively. A nonlinear function $g(\cdot)$ is applied to this value (see Equation (11)). Gates are a unique structure of the LSTM that are can to (optionally) let information through. A gate is made up of a sigmoid activation function $\sigma(\cdot)$ and a point-wise multiplication operator. Furthermore, the output of the gates are nonlinear transforms of the input $\mathbf{X}_t$, recurrent feedback $\mathbf{h}_{t-1}$, and the cell state feedback $\mathbf{C}_t/\mathbf{C}_{t-1}$ (Equation

TABLE I
Mathematical notations.

Symbol	Meaning
w	Sequence length
$\mathbf{x}_i^d$	ith element in the dth sequence
$\mathbf{X}^d$	The sequence $\{\mathbf{x}_1^d, \ldots, \mathbf{x}_w^d\}$
$\mathbf{X}_t$	LSTM input vector at time t
y^d	class label for the dth sequence
$\widehat{y}^d$	RNN prediction for the dth sequence
Δt	Sampling period
$\mathbf{W}, \mathbf{H}, \mathbf{Q}$	RNN weight matrices
PPL	Perplexity

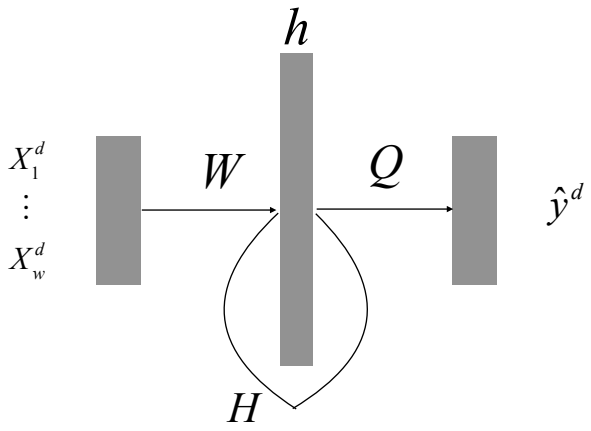


Fig. 1. Visualization of a generic recurrent neural network. $\mathbf{W}, \mathbf{H}, \mathbf{Q}$ represent the weight matrices and h presents a hidden layer of the network, the output label $\widehat{y}^d$ depends on the dependencies between each elements of input sequence $\{\mathbf{x}_1^d, \ldots, \mathbf{x}_w^d\}$, which are captured by the recurrent connection (the circle at the middle of the figure).

(12)-(13) and (15)). Given that the output of the sigmoid is between zero and one then nothing passing through the gate would correspond to a value of zero. Thus, the input gate $\mathbf{i}_t$ determines the parts of scaled input $\mathbf{z}_t$ that will be stored in the new cell state $\mathbf{C}_t$, and the forget gate $\mathbf{f}_t$ determines what parts of old information in the last cell state $\mathbf{C}_{t-1}$ should be removed. It is the input and forget gates together that update the current cell state $\mathbf{C}_t$ (see Equation (14)). The final output of the LSTM cell is a function of $\mathbf{C}_t$, $\mathbf{o}_t$ and an output activation function $h(\cdot)$ (see Equations (15)-(16)). Our implementation uses a vanilla LSTM recurrent neural network, whose memory block consists of a single memory cell, three gates, activation functions, and peephole connections (see Figure 2). The LSTM computes the hidden output sequence $\mathbf{h}_t$ at time t by following iterations:

Block input:

$$\mathbf{z}_t = g(\mathbf{W}_z\mathbf{X}_t + \mathbf{H}_z\mathbf{h}_{t-1} + \mathbf{b}_z) \tag{11}$$

Input gate:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i\mathbf{X}_t + \mathbf{H}_i\mathbf{h}_{t-1} + \mathbf{H}_{ci} \odot \mathbf{C}_{t-1} + \mathbf{b}_i) \tag{12}$$

where $\odot$ denotes a pointwise multiplication.

Forget gate:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f\mathbf{X}_t + \mathbf{H}_f\mathbf{h}_{t-1} + \mathbf{H}_{cf} \odot \mathbf{C}_{t-1} + \mathbf{b}_f) \tag{13}$$

Memory cell:

$$\mathbf{C}_t = \mathbf{z}_t \odot \mathbf{i}_t + \mathbf{f}_t \odot \mathbf{C}_{t-1} \tag{14}$$

Output gate:

$$\mathbf{o}_t = \sigma(\mathbf{W}_o\mathbf{X}_t + \mathbf{H}_o\mathbf{h}_{t-1} + \mathbf{H}_{co} \odot \mathbf{C}_t + \mathbf{b}_o) \tag{15}$$

Block output:

$$\mathbf{h}_t = \mathbf{o}_t \odot h(\mathbf{C}_t) \tag{16}$$

In this classification problem, the class label of the input sequence is predicted after the last element of sequence was processed. For example, the label $\widehat{\mathbf{y}}^d$ of dth sequence $\{\mathbf{x}_1^d, \ldots, \mathbf{x}_t^d\}$ was generated from the last hidden output $\mathbf{h}_t^d$ by:

$$\widehat{\mathbf{y}}^d = \mathbf{Q}\mathbf{h}_t^d + \mathbf{b}_y \tag{17}$$

where the $\mathbf{b}_y$ term denotes bias vector. The input and output activation functions $g(\cdot)$ and $h(\cdot)$ are tangent functions give by:

$$g(x) = h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{18}$$

Finally, the recurrent activation function $\sigma(\cdot)$ is a logistic sigmoid function:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{19}$$

This memory cell also allows LSTM RNN to avoid the vanishing error problem when trained by Backpropagation Through Time (BPTT) algorithm [24]. One of the key advantages of the LSTM RNNs modeling the behavior of the gliomas is that the LSTM can remember dependencies within the sequence to determine the set of PDEs that the tumor is categorized. We show in the experimental results that the LSTM can capture these dependencies.

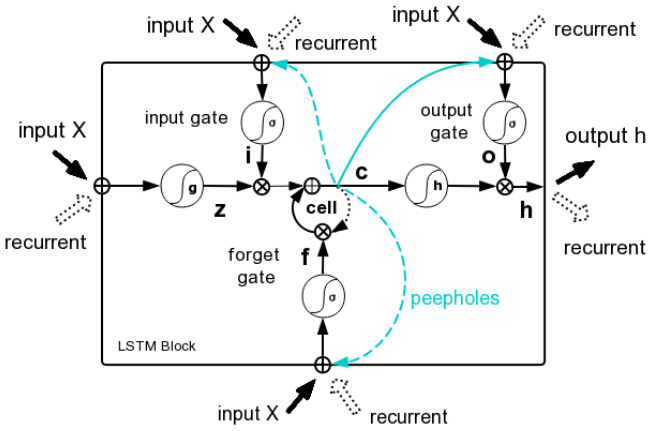


Fig. 2. Visualization of one LSTM memory block. As described in Equation 11-16, $\mathbf{X}$ denotes the input vector of LSTM network at the current time, the recurrent inputs from last time step are presented by the dashed arrow. $\mathbf{z}$ is the scaled block input, $\mathbf{c}$ is the cell state, and $\mathbf{h}$ the output of the block. $\mathbf{i}$, $\mathbf{o}$, and $\mathbf{f}$ are the output value of input gate, output gate and forget gate, respectively. $g(\cdot)$, $h(\cdot)$, and $\sigma(\cdot)$ indicate input, output and sigmoid activation function, respectively. The cyan lines indicate the peephole connections between gates, with the dash lines recurrently connect back to input gate and forget gate at next time step, and solid line connects to output gate at current time step.

B. Data Preprocessing and Conditioning

We originally generated 5000 curves that model the brain tumor growth, as described in Equations (1)-(9). We then selected 20 and 100 curves, such that they are relatively evenly spaced (Figure 3). Figure 3a shows an example of the 20 curves that were selected and Table II shows the PDE parameters for the different experiments.

The time step of the original simulations, $\Delta t = 2.18$ hrs, is too small to reflect clinical reality. For example, an LSTM that is trained to model $\Delta t = 2.18$ hrs means that the medical practitioner is sampling the patient's growth every 2 hrs, which is not realistic. Therefore, we downsampled to $\Delta t = 30$ days and $\Delta t = 60$ days (*i.e,* one and two months, respectively).

We consider the simulation curves as the ground truth. Because real patient data typically includes noise from a variety of sources, like variations in tumor segmentation and quality of images, we add zero-mean Gaussian noise ($\mathcal{N}(0, 0.01)$) to the input signal while keeping the output signal the same. This procedure prevents overfitting and loss of generalization and makes our modeling more stochastic. This type of stochastic addition to the learning problem has been shown to be quite successful with denoising autoencoders [25]. Figure 3 shows the 20/100 curve experiments with the noise added to the signals.

Next, we consider the number of measurements or the number of elements in the sequence. For training, we used

TABLE II
PARAMETERIZATION OF THE PDES FOR EXPERIMENTS USING A DIFFERENT NUMBER OF CURVES WITH THE LSTM MODEL.

	20 Curve Experiments
angiogenesis rate	0.001
mitotic rate	0.25
motility rate	range $(1-100, \text{step} = 5) \times 0.0025$
	100 Curve Experiments
angiogenesis rate	0.001
mitotic rate	0.25
motility rate	range $(1-100, \text{step} = 1) \times 0.0025$

a sequence of 20 measurements. The sequences are randomly partitioned into an 80/20 split for training/testing. To address the question of the optimal number of samples needed to achieve high accuracy, we varied the number of measurements in the testing sequences (w).

We evaluate the performance of the LSTM on different length testing sequences: $w = 2$ to $w = 20$ in two-step increments. This practice is applied to the 20- and 100-curve experiments. The classification label of each curve was assigned based on one-hot coding. Note that a sequence on the dth curve of length w, given by $\{\mathbf{x}_1^d, \ldots, \mathbf{x}_{w-1}^d, \mathbf{x}_w^d\}$, has the single label y^d since all of the samples lie on the same curve.

To address the problem of detecting a change in tumor grade, we experiment with abrupt transitions from one curve to another on a sequence of mass measurements. A piecewise curve test set is designed to mimic the problem where tumor growth trends would change from one curve to another. To synthesize this, we first divide the data into subintervals, then randomly choose one piece of a curve from each subinterval, as shown in Figure 4(b). The design of piecewise growth curves allows us to evaluate the performance when growth trends transition from one curve to another. We considered 20 measurements for each piecewise curve, where the first ten samples belong to a curve d, while the last ten samples belong to another curve d'.

C. Figures of Merit

There are two figures of merit that we use to base the performance of the LSTM for our task. First, we consider the accuracy of the LSTM to identify the correct curve for a sequential prediction task. Nevertheless, because both the parameter space and the number of curves are infinite, we need another test of the closeness of the approximation. Therefore, we consider the perplexity, which is formally given by

$$\text{PPL} = 2^{-\frac{1}{N}\sum \log_2 q_{y_d}(\mathbf{X}^d)}, \tag{20}$$

where N is the number of sequences and $q_{y_d}(\mathbf{X}^d)$ is the probability output of the LSTM for the correct class y_d given a sequence $\mathbf{X}^d$. Note that y_d is *not* the class that the LSTM outputs.

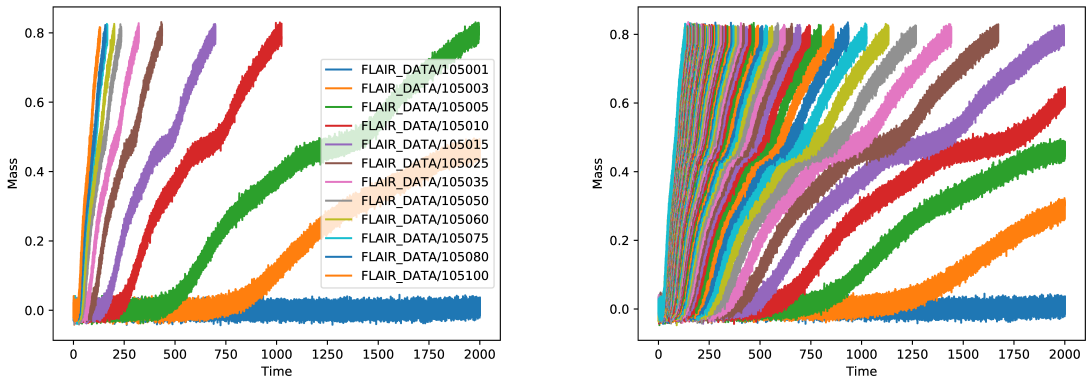
The perplexity better captures the LSTM's ability to report a high probability for the correct class. Perplexity is often applied for language modeling tasks, where there could be a large number of classes [26], [27]. Therefore, an incorrect classification, while incorrect, could still be close to the correct output. For example, the LSTM could predict the wrong curve since our model will choose the class with the largest estimated posterior probability; however, the curve it predicted is the next one below the correct curve, and the correct curve was also predicted with a high posterior probability (though not the largest). Since we expect a large number of classes moving forward, we have decided to include perplexity as a figure of merit.

D. Experimental Results

Our experiments were implemented using Keras [28] with a Tensorflow back-end [29], and run on a NVidia GeForce GTX 950. The LSTM is configured with $\mathbf{W} \in \mathbb{R}^{1\times 50}$, $\mathbf{H} \in \mathbb{R}^{50\times 50}$ and $\mathbf{Q} \in \mathbb{R}^{50\times 20}$ for 20 curves, $\mathbf{Q} \in \mathbb{R}^{50\times 100}$ for 100 curves. The hidden-layer nodes were selected via a small grid search. In our experiments, we observed a limit in the perplexity and accuracy of the LSTM after 50 nodes. Therefore, we only report the figures of merit for an LSTM with 50 nodes. The training window size (i.e., length of the training sequence) was set to 20 for all experiments after some experimental consideration about the physical interpretation between each sample (e.g., one month between samples). Finally, the experiments were run using ten different random seeds, then averaged to smoothen the randomness in the experiments.

Figure 5 shows the predictive accuracy for LSTMs that are trained on sequences sampled every 30 days and 60 days intervals with some measurements ranging from 2 to 20. First, we observe that the predictive accuracy of the two-months samplings is easier to classify. This should not come as a surprise since with Δt^{60} there is a larger change in the values of the curve, thus rendering the classification easier. This observation is true for both the 20- and 100-curves simulations. A second observation is that the accuracy is an increasing function of the number of measurements, w. It takes the LSTM approximately 11 samples to distinguish between the patients' curves. This is observed in both the accuracy and perplexity tests (see Figure 5-6).

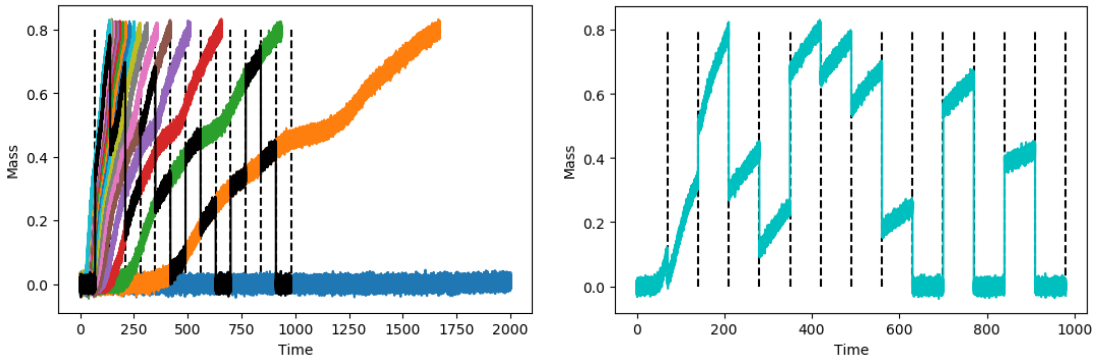
Finally, we study the problem of change in tumor grade by considering piecewise growth curves, where curves d and d' are randomly sampled (see Figure 7). The experiment is designed to simulate a clinical case, whereby the treating physician is monitoring the tumor size on serial scans then tumor undergoes molecular events that lead to a change in grade after the 10th measurement of tumor size (Figures 4(b),



(a) 20 curves

(b) 100 curves

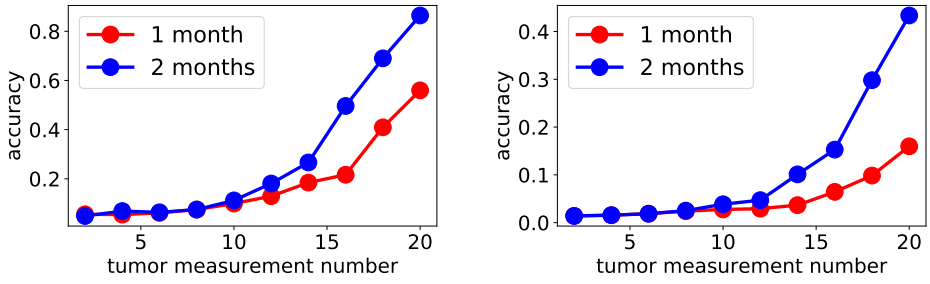
Fig. 3. Two different experiments that use a different number of tumor growth curves. Figure 3(a) has 20 curves that are generated as described in Section III-B. The curves a chosen such that they are approximately evenly spaced. Figure 3(b) shows a similar experiment, but with 100 curves.



(a) Randomly choose one curve from each interval

(b) Detail of piecewise curve, 10 measurements downsampled from each interval

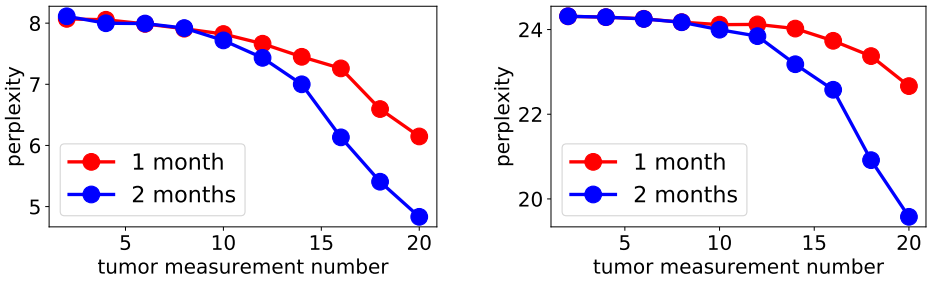
Fig. 4. Two example sequences that show a continuous transition in tumor growth and an abrupt series of changes in the tumor growth. The x-axis is the time step and the y-axis is the mass of the brain tumor. Each curve represents a different set of PDE parameters. The shaded region of Figure 4(a) shows the regions of the curve that are used for testing the LSTM.



(a) 20 curves

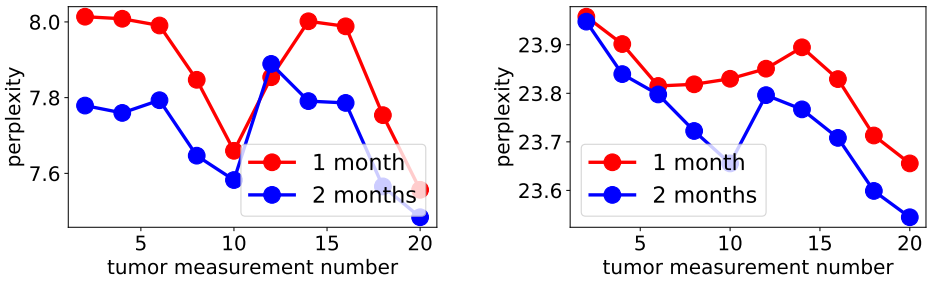
(b) 100 curves

Fig. 5. Accuracy of the LSTM as a function of the number of tumor measurements (window size). The red and blue curves represent simulations where the MRI measurements are obtained at intervals of 1 and 2 months, respectively.



(a) 20 curves (b) 100 curves

Fig. 6. Perplexity of the LSTM as a function of the number of tumor measurements (window size). The red and blue curves represent simulations where the MRI measurements are obtained at intervals of 1 and 2 months, respectively.



(a) 20 curves (b) 100 curves

Fig. 7. Perplexity of the LSTM as a function of number of tumor measurements. The network is tested on piecewise curves and the experiment is designed such that the shift of the growth occurs after the 10th measurement of tumor size (Figure 4(b)). The results shown in (a) and (b) reveal an abrupt increase in the perplexity of the PDEs at sample 12. The red and blue curves represent simulations where the MRI measurements are obtained at intervals of 1 and 2 months, respectively.

and 7). We hypothesize that perplexity can be applied to detect the change in tumor grade, *i.e.*, growth potential (growth curves).

We observe a sharp increase in the perplexity at tumor measurements 12 due to the change in the growth curves as shown in Figure 7. Furthermore, we know from the previous experiment that the accuracy and perplexity are expected to improve over time if all of the samples belong to the same curve; thus demonstrating the promise of LSTMs to solve the parameterization problem for the glioma model.

IV. Conclusion

In this work, we addressed two problems that are of interest to clinical studies that involve malignant brain tumors. First, understanding and modeling the classification and prediction of future behavior of a tumor as it evolves. Second, clinicians are interested in the early detection of dedifferentiation to a higher grade or more aggressive growth. This work addresses these problems by using an LSTM recurrent neural network to solve an inverse problem of parameter estimation for a PDE that describes the tumor. Our results highlight the usefulness of RNN in estimating the parameters of nonlinear PDEs. Savings in computational time are significant because results can be obtained instantaneously once the model is trained. By estimating the parameters of a glioma model, LTSM has the potential of 1) classifying low-grade gliomas by predicting future behavior and 2) detecting a change in tumor growth through the measure of perplexity. We identified perplexity as a figure of merit because of a large number of classes (*i.e.,* the number of PDE parameters that can collect from simulations). Perplexity will become a more meaningful measure of success moving forward since it is a better measurement for a large number of classes, which can become quite large in the context of the inverse problem for tumor modeling. Our results underline the potential use of RNN in solving clinical problems and in aiding clinicians.

Our future work includes incorporating a change detector based on a statistical hypothesis test that checks for a significant decrease in the performance of LSTM [30]–[32]. Furthermore, we will also pursue change detection based on changes of the uncertainty of the LSTMs output.

Acknowledgement

This work was supported by the National Science Foundation under Award DUE-1610911.

References

[1] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues, "The 2007 WHO classification of tumours of the central nervous system," *Acta Neuropathologica*, vol. 114, no. 2, pp. 97–109, 2007.

[2] L. M. Shelton, P. Mukherjee, L. C. Huysentruyt, I. Urits, J. A. Rosenberg, and T. N. Seyfried, "A novel pre-clinical in vivo mouse model for malignant brain tumor growth and invasion," *Journal of Neuro-Oncology*, vol. 99, no. 2, pp. 165–176, 2010.

[3] M. Candolfi, J. F. Curtin, W. S. Nichols, A. G. Muhammad, G. D. King, G. E. Pluhar, E. A. McNiel, J. R. Ohlfest, A. B. Freese, P. F. Moore, J. Lerner, P. R. Lowenstein, and M. G. Castro, "Intracranial glioblastoma models in preclinical neuro-oncology: neuropathological characterization and tumor progression," *Journal of Neuro-Oncology*, vol. 85, no. 2, pp. 133–148, 2007.

[4] Z. Tang, L. Araysi, and H. M. Fathallah-Shaykh, "c-src and neural wiskott-aldrich syndrome protein (n-wasp) promote low oxygen-induced accelerated brain invasion by gliomas," *PLoS One*, vol. 8, no. 9, p. e75436, 2013.

[5] E. Scribner, O. Saut, P. Province, A. Bag, T. Colin, , and H. M. Fathallah-Shaykh, "Glioblastoma grows during anti-angiogenesis: Model to clinical predictions," *PLoS ONE*, vol. 9, no. 12, p. e115018, 2014.

[6] F. Raman, E. Scribner, O. Saut, C. Wenger, T. Colin, and H. M. Fathallah-Shaykh, "Computational Trials: Unraveling Motility Phenotypes, Progression Patterns, and Treatment Options for Glioblastoma Multiforme," *PLoS ONE*, vol. 11, no. 1, p. e0146617, 2016.

[7] E. Scribner, J. R. Hackney, H. C. Machemehl, R. Afiouni, K. R. Patel, and H. M. Fathallah-Shaykh, "Key rates for the grades and transformation ability of glioma: model simulations and clinical cases," *Journal of Neuro-Oncology*, vol. 133, no. 2, pp. 377–388, 2017.

[8] Z. Luo, J. Zhu, R. Wang, and I. M. Navon, "Proper orthogonal decomposition approach and error estimation of mixed finite element methods for the tropical pacific ocean reduced gravity model," *Computer Methods in Applied Mechanics and Engineering*, vol. 194, no. 41–44, pp. 4184–4195, 1996.

[9] L. Sirovich, "Turbulence and the dynamics of coherent structures. i - coherent structures. ii - symmetries and transformations. iii - dynamics and scaling," *Quarterly of Applied Mathematics*, vol. 45, pp. 561–571, 1987.

[10] L. Sirovich and J. D. Rodriguez, "Coherent structures and chaos: A model problem," *Physics Letters A*, vol. 120, no. 5, pp. 211–214, 1987.

[11] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America A*, vol. 4, no. 3, pp. 519–524, 1987.

[12] A. Rehm, E. Scribner, and H. M. Fathallah-Shaykh, "Proper orthogonal decomposition for parameter estimation in oscillating biological networks," *Journal of Computational and Applied Mathematics*, vol. 258, pp. 135–150, 2014.

[13] M. Boulakia, E. Schenone, and J. F. Gerbeau, "Reduced-order modeling for cardiac electrophysiology. application to parameter identification," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 28, pp. 727–744, 2012.

[14] A. R. M. Roa and K. Lakshmi, "Parameter estimation technique combining pod with hybrid adaptive swarm intelligence algorithm," in *International Conference on Advances in Civil Engineering*, pp. 57–63, 2011.

[15] M. Kahlbacher and S. Volkwein, "Estimation of regularization parameters in elliptic optimal control problems by pod model reduction," in *System Modeling and Optimization* (A. Korytowski, K. Malanowski, W. Mitkowski, and M. Szymkat, eds.), vol. 312 of *IFIP Advances in Information and Communication Technology*, pp. 307–318, Springer Boston, 2009.

[16] H. V. Ly and H. T. Tran, "Proper orthogonal decomposition for flow calculations and optimal control in a horizontal cvd reactor," tech. rep., 1998.

[17] Y. H. Hu and J.-N. Hwang, eds., *Handbook of Neural Network Signal Processing*. CRC Press, 1st edition ed., 2001.

[18] J. R. Raol, "Parameter estimation of state space models by recurrent neural networks," *IEE Proceedings - Control Theory and Applications*, vol. 142, no. 2, pp. 114–118, 1995.

[19] J. Raol and H. Madhuranath, "Neural network architectures for parameter estimation of dynamical systems," *IEE Proceedings - Control Theory and Applications*, vol. 143, no. 4, pp. 387–394, 1996.

[20] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *Journal of the American Medical Informatics Association*, vol. 24, p. 361370, March 2017.

[21] J. Zhou, X. Hong, F. Su, and G. Zhao, "Recurrent convolutional neural network regression for continuous pain intensity estimation in video," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1535 – 1543, July 2016.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[23] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," in *A Field Guide to Dynamical Recurrent Neural Networks* (I. Press, ed.), 2001.

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[25] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *International Conference on Machine Learning*, 2008.

[26] S. Kombrink, T. Mikolov, M. Karafiát, and L. Burget, "Recurrent neural network based language modeling in meeting recognition," in *Interspeech*, 2011.

[27] T. Mikolov, S. Kombrink, L. Burget, J. Černocký, and S. Khudanpur, "Extensions of recurrent neural network language model," in *International Conference on Acoustics, Speech and Signal Processing*, pp. 5528–5531, 2011.

[28] F. Chollet *et al.*, "Keras." https://github.com/keras-team/keras, 2015.

[29] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015.

[30] L. Wasserman, *All of Statistics: A concise course in Statistical Inference*. Springer, 2005.

[31] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, "Adaptive strategies for learning in nonstationary environments: a survey," *Computational Intelligence Magazine*, vol. 10, no. 4, pp. 12–25, 2015.

[32] M. Baena-Garcia, J. del Campo-Avila, R. Fidalgo, A. Bifet, R. Gavaldua, and R. Morales-Bueno, "Early drift detection method," in *International Workshop on Knowledge Discovery from Data Streams*, 2006.