

基于机器学习方法的超声导波结构健康监测研究

高东岳

(中山大学航空航天学院, 广州 510275)

摘要 随着复合材料结构健康监测技术的广泛应用, 基于先进传感与数据挖掘技术的结构损伤识别已成为国内外的研究热点问题。本文以机器学习理论技术为基础, 对复合材料结构受拉伸疲劳载荷造成的损伤进行定量化识别分析和研究, 针对结构疲劳损伤监测数据之间存在的时序相关性, 构建出基于随机森林的结构损伤识别模型, 为复合材料结构健康监测分析提供了新的思路和方法。主要研究内容如下: 提取不同路径的信号特征信息, 包括信号能量, 信号峰值, 信号主峰到达时间。选择 MSE 作为不纯度函数, 设计并构建了一种基于随机森林的结构损伤识别模型; 为了验证本文所提出的基于机器学习的复合材料结构损伤识别模型的有效性, 以 Bookshelf 框架结构数据为基准进行对比试验分析。结果表明: 基于信号能量, 信号峰值, 信号主峰到达时间特征的随机森林结构损伤识别模型可以有效识别与预测损伤。本文试验数据来自 NASA 与斯坦福大学进行的联合损伤表征试验。

关键词 结构健康监测; 超声导波; 机器学习; 随机森林回归; 损伤监测

Ultrasonic Guided Wave Structural Health Monitoring Technology Based on Machine Learning Method

GAO Dongyue

(Sun Yat-Sen University, School of Aeronautics and Astronautics, Guangzhou 510275)

ABSTRACT With the application of the composite structural health monitoring technology, structural damage identification technology based on advanced sensor and data mining was widely researched. In this paper, based on the theory of machine learning technology, structure of composite material by the tensile fatigue load through the quantitative identification analysis and research on the damage for structural fatigue damage monitoring the temporal correlation of data, constructed the model of structural damage identification based on random forest, for the analysis of composite structure health monitoring provides a new thought and method. The main research contents are as follows: Propose signal characteristic information of different paths, including signal energy, signal peak, and the arrival time of the main peak of the signal; Select MSE as the impurity function, and design and construct a random forest; In order to validate the presented composite structure damage identification model based on machine learning effectiveness, Bookshelf frame structure data as benchmark to compare the experimental analysis results show that the energy of the signal, the signal peak value, peak arrival time signal characteristics of random forest structure damage identification model can effectively identify and predict damage in this paper, the experimental data of joint damage characterization were extracted from NASA and Stanford university's experiment.

KEYWORDS health monitoring; ultrasonic guided wave; machine learning method; random forest structure; damage monitoring

1 引言

复合材料结构的大量应用是实现航天运载器结构减重增效任务的坚实基础。航天运载器的可重复

使用需求, 以及复杂恶劣的载荷环境, 使得其结构和以往型号相比, 可能面临更多的潜在风险。结构健康管理技术可以有效地监测结构损伤的发生与扩

展,并预测损伤发展趋势,为航天复合材料结构的服役安全性和可靠性提供保障。对于大面积板壳类复合材料结构,基于 Lamb 波信号特征分析方法可以实现结构内部裂纹与脱粘损伤的定位与量化识别,被证明是有效的复合材料完整性监测技术。针对结构损伤预测需求,研究人员基于 Lamb 波信号特征与机器学习方法,开展了损伤扩展预测技术的研究。该技术以不同状态下 Lamb 波信号特征组成样本库,以无损检测手段获得的损伤尺寸或类型为标签,训练模型,优化参数实现对损伤未来扩展趋势的预测或对损伤形成过程进行回溯。

机器学习 (ML) 方法属于数据驱动技术,常用于存在不确定性的复杂非线性现象建模中,基于以往数据进行训练,在缺乏物理模型的情况下,这些非参数方法提供了成功的损伤检测和估计。人工神经网络方法具有建模复杂非线性函数的能力,Su 和 Ye^[1]提出了一种基于前馈神经网络的复合材料层合板损伤检测方法。类似的多层前馈神经网络也已应用于厚钢梁的损伤检测中^[2]。Dworakowski, Z.^[3]和 Sbarufatti, C.^[4]将基于损伤指数的复杂几何结构兰姆波响应与神经网络相结合进行损伤检测。也有人尝试将特征提取技术与小波神经网络等结构中的学习集成起来。

在训练数据缺乏的情况下,支持向量机也具有较的学习能力和泛化能力,因此很多研究倾向于利用支持向量机进行结构损伤分类识别。各种非线性核函数被用于支持向量机的非线性建模,并提出了一些适用于特定损伤类型的核函数。Agarwal 和 Mitra^[5]报道了基于兰姆波的铝板损伤检测方法,该方法使用 ANN 和 SVM 方法,并结合匹配追踪来增加数据的稀疏性和去噪。比较了两层前馈神经网络和三次多项式核函数支持向量机的结果。如前所述,利用匹配追踪处理输入兰姆波响应,几种概率方法也适用于损伤检测和估计^[6]。贝叶斯分类器和概率图形模型使用证据数据进行预测和估计。Ng, C.T.^[7]提出了基于导波响应的类梁结构损伤定量检测的贝叶斯方法。它还被用于基于小波变换获得的飞行信息时间的兰姆波损伤检测^[8]。也可以使用 k 近邻方法等非参数技术。先进的统计模型,如广义回归和逻辑回归,也可以用于 Lamb 波的损伤检测。

目前,基于 Lamb 波的损伤识别方法的发展趋

势是利用大量分布式传感器覆盖整个被测物体表面,按照“激发器→接收传感器”的规则组成很多个信号传播路径,同时选择多个频率激发传感器,以求最大限度的掌握结构内部信息。多路径、多频率意味着多信号特征,面临着越来越高的样本数据维度,一些机器学习方法出现了泛化能力弱,过拟合的问题。随机森林回归方法能够处理很高维度的数据;随机选择特征子集并可以分析不同特征的重要性;在创建随机森林时,对泛化误差使用的是无偏估计,模型泛化能力强;该方法不需要单独划分交叉验证集,在训练过程中,树与树之间是相互独立的,训练速度快;对样本缺失值不敏感,即使有很大一部分的特征遗失,仍可以维持准确度。

本文以 NASA 与斯坦福大学进行的联合损伤表征试验数据为爬取对象,利用多个路径不同频率下 Lamb 波信号与 X 射线成像损伤尺寸数据,分析信号能量、到达时间、信号峰值等信号特征与损伤尺寸的相关性,采用 RF 回归算法建立每个损伤尺寸反演的随机森林回归模型。通过对模型的测试的处理,认为该模型预测损伤尺寸较其他算法精度更高,泛化能力更强,对实现复合材料损伤识别与预测需求帮助很大。

2 随机森林回归算法

随机森林属于 Bagging 类算法,大致思路是训练多个弱模型打包起来组成一个强模型,在随机森林回归中,弱模型选用决策树。在训练阶段,随机森林使用 Bootstrap 采样从输入训练数据集中采集多个不同的子训练数据集来依次训练多个不同决策树;在预测阶段,随机森林将内部多个决策树的预测结果取平均得到最终的结果。回归决策子树 T 和特征子空间中的特征个数 f 决定着模型最终的预测性能。

本文实现的随机森林回归模型是由多个二叉决策树打包组合而成的,训练模型便是训练多个二叉决策树。在训练二叉决策树模型的时候需要考虑怎样选择切分变量(来自不同路径的信号特征)、切分点,以及怎样衡量一个切分变量、切分点的好坏。针对于切分变量和切分点的选择,本文采用穷举法,即遍历每个特征和每个特征的所有取值,最后从中找出最好的切分变量和切分点;针对于切分变量和切分点的好坏,一般以切分后节点的不纯度来衡量,

即各个子节点不纯度的加权。

$$G(x_i, v_{ij}) = \frac{n_{left}}{N_s} H(X_{left}) + \frac{n_{right}}{N_s} H(X_{right}) \quad (1)$$

式中, x_i 为某一个切分变量, v_{ij} 为切分变量的一个切分值, n_{left} 、 n_{right} 、 N_s 分别为切分后左子节点的训练样本个数、右子节点的训练样本个数以及当前节点所有训练样本个数, X_{left} 、 X_{right} 分为左右子节点的训练样本集合, $H(X)$ 为衡量节点不纯度的函数, 分类和回归任务一般采用不同的不纯度函数, 本文选择 MSE 作为不纯度函数, 其中,

$$H(X_m) = \frac{1}{N_m} \sum_{i \in N_m} (y - y_m)^2, y_m \text{ 为当前节点样本目标变量的平均值。}$$

决策树中某一节点的训练过程在数学等价于下面优化问题:

$$(x^*, v^*) = \arg \min_{x, v} G(x_i, v_{ij}) \quad (2)$$

即寻找 G 最小的切分变量和切分点。本文中,

$$H(X_m) = \frac{1}{N_m} \sum_{i \in N_m} (y - y_m)^2 \text{ 即针对一个切分点:}$$

$$G(x, v) = \frac{1}{N_s} \left(\sum_{y_i \in X_{left}} (y_i - \bar{y}_{left})^2 + \sum_{y_i \in X_{right}} (y_i - \bar{y}_{right})^2 \right)。$$

在建模过程中, 首先在 Bagging 思想的指导下随机产生样本子集; 然后基于随机子空间思想, 随机抽取 f 个特征, 进行节点分裂, 构建单棵回归决策子树; 接下来重复步骤 1、步骤 2, 构建 T 棵回归决策子树, 每棵树自由生长, 形成随机森林; 棵决策子树的预测值取平均, 作为最终结果; 重复以上步骤, 直到 OOB 误差最低。

3 加载试验

本文爬取的数据是斯坦福结构与复合材料试验室 (SACL) 与美国宇航局艾姆斯研究中心的预测卓越中心 (PCoE) 合作对 CFRP 复合材料进行的疲劳老化测试试验所公开的数据, 包括损伤尺寸数据, Lamb 波数据和载荷大小数据。在此试验中, 一组试件在频率为 5.0 Hz, 应力比为 $R=0.14$ 的循环载荷控制回路下进行拉伸疲劳试验, 试件如图 1 所示。

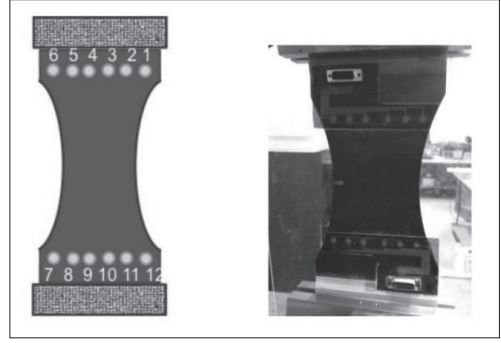


图 1 试件实物图

此试验共测试了三种不同铺层形式的复合材料标准试件, 其铺层形式分别为:

- (1) Layup 1: $[0_2/90_4]_s$;
- (2) Layup 2: $[0/90_2/45/-45/90]$;
- (3) Layup 3: $[90_2/45/-45]_2$ 。

本文取 Layup 2 作为数据爬取对象。

试件采用 Torayca T700G 单向碳预浸料制作, 尺寸为 $15.24\text{cm} \times 25.4\text{cm}$, 狗骨几何形状, 在试件中部设计宽度为 5.08mm, 长度为 19.3mm 的缺口, 用以诱导应力集中。所有试验均在 MTS 拉伸/疲劳试验机上按照 ASTM 标准 D3039 和 D3479 进行。疲劳循环测试每 5 万次就停止一次, 以收集所有路径和频率的 PZT 传感器数据以及样品的 x 射线图像。

负责激发与接收 Lamb 波的传感器网络为粘接在每个样品表面的两套压电晶片传感器 SMART Layer (Acellent Technologies, Inc. 公司出品, 共 12 个传感器)。该套传感器网络共有 6 个激发器和 6 个接收传感器用以监控样件的 Lamb 波传播特性。每一个 36 “激发器→接收器路径” 拥有从 150KHz 到 450 KHz 的 7 个不同的驱动频率, 平均输入电压为 50V, 增益为 20 dB。利用 L2S17 试件上的传感器网络, 在不同疲劳循环次数条件下, 激发 250KHz 的 Lamb 波, 采集路径 6#→7# 的传感器信号如图 2 所示。

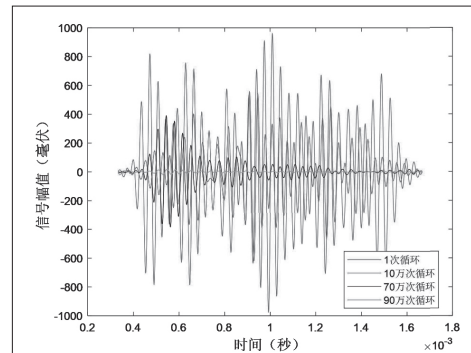
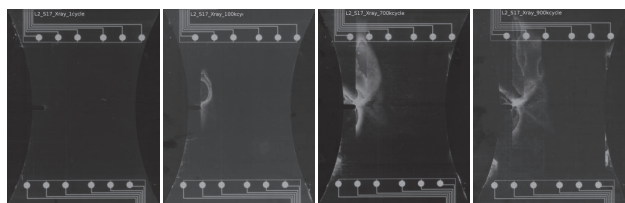


图 2 不同疲劳循环次数下路径 6#→7#, 250KHz 的 Lamb 波信号

图2 L2S17试件为不同疲劳循环次数下路径6#→7#, 250KHz的Lamb波信号。由图可见, 随着载荷循环次数的增加, Lamb波载体的边界条件与材料属性逐渐被破坏, 传感器信号逐渐变化, 能量逐渐衰减, 10万次循环与1次循环相比, 信号能量从S0模态转化为A0模态, 最后信号衰减为零。

经过1、10万、70万、90万次循环疲劳加载, 损伤的具体扩展趋势, 可由试件的X射线图像显示出来, 如图3所示。



(a) 1次循环 (b) 10万次循环 (c) 70万次循环 (d) 90万次循环

图3 不同循环疲劳次数下试件的X射线成像图

可以清晰地看出随着循环次数的增长, 损伤在

不断的扩张, 直至结构失效。70万次循环的损伤图像清晰地显示, 结构损伤已经向传感器网络内部发展, 90万次循环下损伤已经破坏了传感器下方结构, 所以90万次循环下传感器已经无法有效激发Lamb波信号。试验根据X射线图像测量, 给出了不同循环次数的损伤尺寸, 如图4所示。

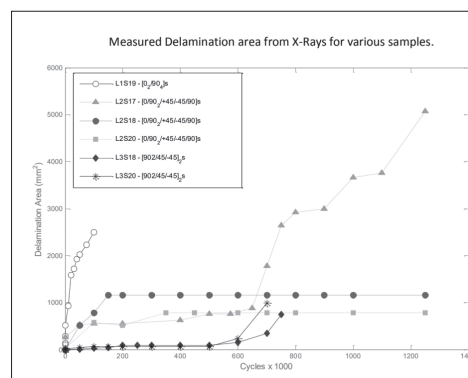


图4 不同循环次数的损伤尺寸

表1 不同疲劳循环次数下的损伤尺寸变化规律

次数(次)	1	1e4	2e4	3e4	4e4	5e4	6e4	7e4	8e4	9e4
尺寸(mm ²)	263.2	312.82	362.44	412.06	461.68	511.3	320.34	377.48	434.62	491.76
次数(次)	10e4	12.5e4	15e4	17.5e4	2e40	22.5e4	27.5e4	30e4	32.5e4	35e4
尺寸(mm ²)	548.9	548.9	548.9	548.9	548.9	570	612	633	591.15	599.6
次数(次)	40e4	42.5e4	45e4	47.5e4	50e4	52.5e4	55e4	57.5e4	62.5e4	65e4
尺寸(mm ²)	616.5	650.35	684.2	718.05	751.9	658	564	470	257	171.05
次数(次)	67.5e4	7e40	72.5e4	75e4	77.5e4	80e4	85e4	90e4		
尺寸(mm ²)	1250.9	1789.5	1342	894.75	2246	2909.8	2940	2970		

由于超过90万次循环之后, 传感器的激发边界条件被破坏, 所以试验中不再采集Lamb波数据。因此, 本文对数据的爬取只截止到90万次循环。

4 模型设计

在模型设计过程中,算法流程如图5所示。

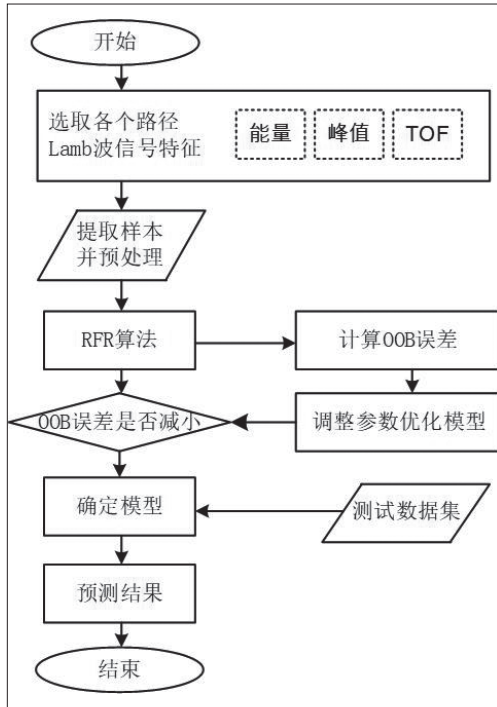


图5 算法流程

步骤1 选取传感器网络中不同路径 ($Path_{i \rightarrow j}$, $i \in 1\# \sim 6\#, j \in 7\# \sim 12\#$) 的信号特征, 收集数据并进行预处理, 得到样本库;

步骤2 将建模数据作为 RFR 算法的输入, 训练得到 RFRP 模型;

步骤3 调整参数, 优化模型, 直到 OOB 误差最低;

步骤4 应用优化后的模型对新的数据进行预测。

特征选取涉及 Lamb 波的 ($1 \sim 6\#$) \rightarrow ($712\#$) 不同路径的 36 组独立信号作为样本特征, 该特征具备损伤尺寸相关、信息量大、差异性特点。经过充分的调研分析, 选取能量、峰值、TOF 分别训练模型。

36 组项因素为特征分析对象。将不同条件下采集的信号作为不同样本进行数据爬取, 整理出

S2L17 的 Lamb 波数据, 共 756 条。随机选取其中的 75% 作为训练集建立模型, 剩余的 25% 用来测试模型的准确率。对原始数据进行归一化预处理, 形成一个 756×36 的矩阵:

$$A = \begin{bmatrix} a_{1f_1} & a_{1f_2} & \dots & a_{1f_{36}} \\ a_{2f_1} & a_{2f_2} & \dots & a_{2f_{36}} \\ \vdots & \vdots & \ddots & \vdots \\ a_{756f_1} & a_{756f_2} & \dots & a_{756f_{36}} \end{bmatrix} \quad (3)$$

由于 RFR 算法对数据的单位和量纲并不敏感以及两大随机特性, 所以不需要对整理好的数据进行归一化处理和特征选择。原始数据经过预处理后, 得到矩阵 A, 直接作为 RFR 算法的输入。利用 Bootstrap 重采样随机生成样本子集, 同时产生 OOB 数据。在树的每个节点随机选择 f 个特征进行分裂, 建立多个损伤尺寸预测模型, 并利用 OOB 数据计算 OOB 误差。最后将多个模型的预测值取平均作为模型输出。

采取通用的模型精度和效率作为度量指标, 进行 RFRP 模型的参数调整 and 性能分析。包括 MRE (mean relative error, 平均相对误差) 和确定性系数 (R^2)。其中, MRE 越小, R^2 越大, 说明模型精度越高。

5 结果分析

以拉伸疲劳测试试验数据为训练样本, 以损伤尺寸值为因变量, 选择各个信号路径的能量、峰值、到达时间作为自变量, 采用随机森林回归算法构建 3 个信号特征的损伤尺寸值估算模型。利用 R^2 指标评价的 3 种特征的机器学习模型学习能力和预测能力。

在模型构建中, 根据经验及多次试验, 确定三个特征随机森林回归算法中分类树个数均为 2000, 分割变量均为 3。各信号特征决策树模型预测值与实测值关联性如图 6 所示。

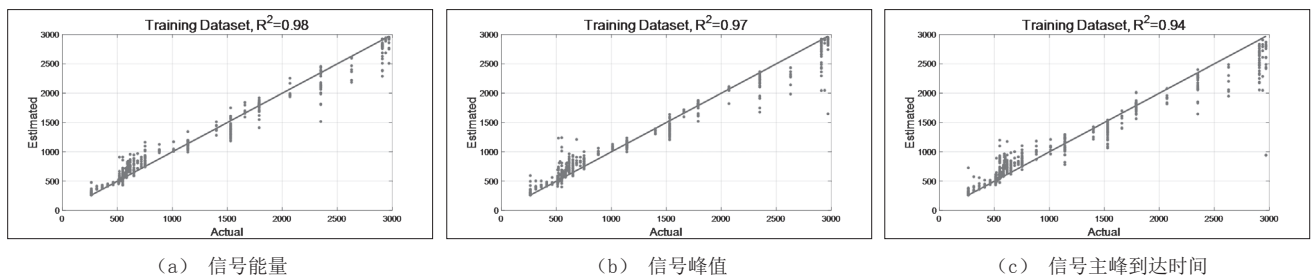
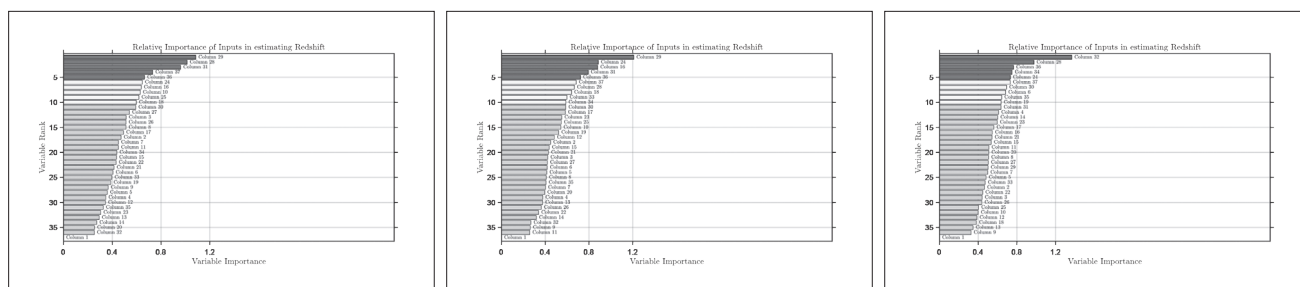


图6 各信号特征决策树模型预测值与实测值关联性

结果表明,由这些模型推算得到的预测值与实测值之间都呈极显著相关, R^2 都较理想,说明由这

些特征监测损伤尺寸值是可行的。信号路径重要性分析如下图 7 所示。



(a) 信号能量

(b) 信号峰值

(c) 信号主峰到达时间

图 7 路径重要性分析

对于信号能量特征,样本 29 (5-10) /28 (5-9) /31 (5-12) /37 (6-12) /36 (6-11) 是重要路径;对于信号峰值特征,样本 29 (5-10) /24 (4-11) /16 (3-9) /31 (5-12) /36 (6-11) 是重要路径;对于信号主峰到达时间特征,32 (6-7) /28 (5-9) /36 (6-11) /34 (6-9) /24 (4-11) 是重要路径。通过分析可知,靠近损伤的路径信号能量与峰值特征更多的反映了损伤的扩展,对于信号主峰到达时间,距离损伤最远的 6→7 路径出现了达到时间最大的变化。这一结论符合物理规律。

6 结语

目前损伤预测模型建立过程中,选择一种合适的信号特征有助于提高损伤尺寸预测的精度。本文利用随机森林算法基于 Lamb 波信号能量反演了复合材料试件在不同疲劳循环次数下损伤尺寸,在一定程度上可以反演疲劳损伤在复合材料内部生长状况。该算法构建的模型具有较高的预测精度,可为复合材料疲劳损伤的无损、快速监测提供一种新的方法。

本研究发现信号能量、信号峰值与信号主峰到达时间都可以作为识别监测损伤的数据特征。但不同信号特征反映损伤的重点路径不同。

参考文献

- [1] Kim, E.J., et al., Investigating electro-mechanical signals from collocated piezoelectric wafers for the reference-free damage diagnosis of a plate. *Smart Material Structures*. 20(6):065001.
- [2] Atashipour, S.A., et al., An effective damage identification approach in thick steel beams based on guided ultrasonic waves for structural health monitoring applications. *Journal of Intelligent Material Systems & Structures*. 24(5):584-597.
- [3] Dworakowski, Z., et al., Application of artificial neural networks for compounding multiple damage indices in Lamb-wave-based damage detection. *Structural Control & Health Monitoring*. 22(1):50-61.
- [4] Sbarufatti, C., G. Manson and K. Worden, A numerically-enhanced machine learning approach to damage diagnosis using a Lamb wave sensing network. *Journal of Sound & Vibration*. 333(19):4499-4525.
- [5] Agarwal, S. and M. Mitra, Lamb wave based automatic damage detection using matching pursuit and machine learning. *Smart Materials & Structures*. 23(8):085012.
- [6] Xie, J., et al. Prediction of Low-velocity Impact Damage in Sandwich Composite Beams. in *Aiaa/asce/ahs/asc Structures, Structural Dynamics, & Materials Conference*. 2015.
- [7] Ng, C.T., On the selection of advanced signal processing techniques for guided wave damage identification using a statistical approach. *Engineering Structures*. 67(MAY 15):50-60.
- [8] Gang, Y., A Bayesian approach for damage localization in plate-like structure using Lamb waves. *Smart Materials & Structures*, 2013. 22(3):035012.